Raffaello Cortina Editore

Cass R. Sunstein

Manipolazione

Cos'è, perché è un problema e come difendersi

Scienza e Idee

Collana fondata da Giulio Giorello

Abbiamo l’illusione di poter scegliere liberamente, ma il gioco è truccato e spesso ci spinge in direzioni che compromettono il nostro benessere, la nostra salute e le nostre democrazie. Cass Sunstein, pioniere della teoria del *nudge*, la “spinta gentile”, ci guida attraverso le nuove forme della manipolazione digitale, che sfrutta tecnologie, algoritmi e AI generativa, e attraverso le minacce poste da deepfake, social media e tecniche di programmazione ingannevoli, che inducono le persone a buttare al vento tempo e denaro.
Basato su decenni di ricerche pionieristiche nel campo della scienza del comportamento, questo libro delinea i passi che possiamo compiere per difenderci e contrastare la manipolazione nella nostra vita quotidiana.

Cass R. Sunstein è Robert Walmsley University Professor presso la Harvard Law School, dove è fondatore e direttore del programma di studi in Economia comportamentale e politiche pubbliche. *Nudge. La spinta gentile* (con R.H. Thaler, Milano, 2009) è un manuale cardine dell'economia comportamentale. Nelle nostre edizioni ha pubblicato *Guardate meglio* (con T. Sharot, 2024) e *Come diventare famosi* (2025).

Scienza e idee

Collana fondata da Giulio Giorello

Consulenza scientifica
Telmo Pievani, Corrado Sinigaglia

Dal catalogo

Cass R. Sunstein
Come diventare famosi
La scienza segreta del successo

Tali Sharot, Cass R. Sunstein
Guardate meglio
Perché l'abitudine ci rende ciechi

Steven Sloman
Il prezzo della certezza
Perché i nostri valori più radicati ci inducono in errore

Gerd Gigerenzer
Perché l'intelligenza umana batte ancora gli algoritmi

Gerd Gigerenzer
Imparare a rischiare
Come prendere decisioni giuste

Peter Burke
Ignoranza
Una storia globale

Olivier Sibony
Stai per commettere un terribile errore!
Come evitare le trappole del pensiero

Cass R. Sunstein

Manipolazione

Cos’è, perché è un problema
e come difendersi

Raffaello Cortina Editore

www.raffaellocortina.it

Titolo originale
Manipulation. What it is, Why it's Bad, What to do About it
© 2025 by Cass R. Sunstein. All rights reserved.

Traduzione
Virginio B. Sala

Copertina
mara scanavino project | Alberto Lameri

ISBN 978-88-328-5908-9
© 2026 Raffaello Cortina Editore
Milano, via Rossini 4

Prima edizione: 2026

INDICE

Prefazione

PARTE PRIMA

Teoria

1. Definire la manipolazione
2. Valutare la manipolazione
3. Il diritto a non essere manipolati

PARTE SECONDA

Applicazioni

4. Deepfake: manipolati dalla tecnologia
5. Scarsità e attenzione: manipolati dal pantano
6. Il problema Barbie: manipolati dalla pressione sociale

PARTE TERZA

Il futuro

7. Pungoli senza manipolazione, 1
8. Pungoli senza manipolazione, 2

9. Intelligenza artificiale: promessa e minaccia

Epilogo: prevenire il furto
Ringraziamenti

A Sendhil Mullainathan

PREFAZIONE

Questo è un libro di cui non potete fare a meno.

Forse credete di capire le altre persone, ma la vostra conoscenza scalfisce appena la superficie. La vera maestria sta nell'arte della manipolazione, una pratica antica quanto l'umanità stessa.

Questo libro è la vostra guida all'interno di quel regno proibito, dove le regole sono riscritte e la posta è la vita stessa. Vi rivelerà le tecniche, la psicologia e le strategie utilizzate dalle personalità più influenti del mondo.

Imparerete non solo a riconoscere la manipolazione, ma anche a metterla in pratica con precisione.

Scorrendo queste pagine, lasciate che la potenza di questa conoscenza penetri in voi. Immaginate come potrebbe cambiare la vostra vita con la capacità di influenzare chiunque, ovunque, in qualsiasi momento.

Questo non è solo un libro: è una chiave per liberare un potenziale che non avevate mai pensato potesse esistere. Ma state attenti: questo viaggio vi trasformerà. Una volta abbracciati quei segreti, non si può più tornare indietro.

Siete pronti a sfruttare il potere che ha forgiato imperi e conquistato legioni? La scelta sta a voi, ma ricordate: a ogni grande potere corrisponde una grande responsabilità. Benvenuti nel mondo della manipolazione.

Lo confesso: non ho scritto io i paragrafi precedenti in corsivo. Ho chiesto a ChatGPT di scrivere un incipit manipolativo per un libro sulla manipolazione, e questo è il risultato.[1] Per certi versi queste frasi mi irritano, ma non sono poi male. Promettono di rivelare segreti. Sono un tentativo di seduzione. Fanno leva sulle emozioni di chi legge ("*Questo non è solo un libro: è una chiave per liberare un potenziale che non avevate mai pensato potesse esistere*").

Ecco un passo più noto.[2] Lo conoscete, vero?

> Il serpente era il più astuto di tutti gli animali selvatici che Dio aveva fatto e disse alla donna: “È vero che Dio ha detto: ‘Non dovete mangiare di alcun albero del giardino?’”.
>
> Rispose la donna al serpente: “Dei frutti degli alberi del giardino noi possiamo mangiare, ma del frutto dell’albero che sta in mezzo al giardino Dio ha detto: ‘Non dovete mangiarne e non lo dovete toccare, altrimenti morirete’”.
>
> Ma il serpente disse alla donna: “Non morirete affatto! Anzi, Dio sa che il giorno in cui voi ne mangiaste si aprirebbero i vostri occhi e sareste come Dio, conoscendo il bene e il male”.
>
> Allora la donna vide che l’albero era buono da mangiare, gradevole agli occhi e desiderabile per acquistare saggezza; prese del suo frutto e ne mangiò, poi ne diede anche al marito, che era con lei, e anch’egli ne mangiò.
>
> Allora si aprirono gli occhi di tutti e due e conobbero di essere nudi; intrecciarono foglie di fico e se ne fecero cinture.

Il serpente era un manipolatore, un seduttore. Conosceva bene la psicologia umana. Sapeva come far leva sulle emozioni dei suoi interlocutori: “i vostri occhi si apriranno, e sarete come Dio”. (Era un esperto di bias ottimistico? Di scienza comportamentale? Ovviamente.) In effetti, ChatGPT ha utilizzato alcune delle stesse tecniche del serpente. Ha promesso di rivelare segreti nascosti e di aprire gli occhi dei lettori.

Viviamo in un’epoca in cui la manipolazione è onnipervasiva, specialmente online. Uno dei suoi scopi fondamentali è attirare l’attenzione. Se una società di social media può catturare la vostra attenzione, può far pagare qualcosa agli inserzionisti pubblicitari. Se un inserzionista riesce a catturare la vostra attenzione, può riuscire a scucirvi un po’ di soldi. I politici vogliono i vostri soldi e i vostri voti, e potrebbero riuscire a manipolarvi e farvi fare quello che vogliono.

L’abilità nel manipolare le persone però va molto al di là del catturare l’attenzione. I manipolatori, spesso aiutati (o proprio costituiti) dall’intelligenza artificiale, possono far leva sui vostri desideri, le vostre speranze, le vostre paure e i vostri punti deboli. Se sanno qualcosa di voi, possono sapere molto su di voi. Possono riuscire a convincervi ad agire come vorrebbero. Vogliono il vostro denaro, e vogliono il vostro tempo.

La manipolazione è una minaccia grave per il buon funzionamento dei mercati, che dipende da scelte informate. La manipolazione è anche una minaccia grave per la democrazia e, soprattutto, compromette la libertà. È una minaccia alla nostra capacità di esercitare un’azione autonoma.

Si è scritto molto sulla coercizione e quasi altrettanto sulla menzogna e l’inganno; al confronto, la manipolazione è un argomento trascurato. Un motivo, forse, è che il termine è difficile da definire; un altro sono forse la sua onnipresenza e la sua inevitabilità. Se accendete un televisore o andate

online, probabilmente vi imbatterete in un tentativo di manipolazione. Se uno dei vostri familiari non tenta di manipolarvi nel corso del prossimo mese, vuol dire che avete una famiglia fuori dal comune. Dato che è difficile riuscire a formularne una definizione chiara, e dato che alcune forme di manipolazione sono inevitabili o innocue, l'argomento presenta sfide inusuali. Qui cercherò di superare qualcuna di quelle sfide. Argomenterò contro la manipolazione (nella maggior parte dei casi), ma metterò anche in evidenza strumenti di influenza che si potrebbe pensare siano manipolativi e malvagi, ma in realtà non sono né una cosa né l'altra. Dobbiamo opporci a molte forme di manipolazione, ma dobbiamo anche sapere che cosa la manipolazione non è.

Il libro è diviso in tre parti. La parte prima prende in considerazione interrogativi sul piano teorico: che cos'è la manipolazione, in che senso è una cosa sbagliata, che cosa si può fare in proposito. Dedica molta attenzione ad autonomia, dignità e benessere sociale. Collega coercizione, menzogna, inganno e manipolazione (i quattro Cavalieri dell'Apocalisse?) e mostra che sollevano un intreccio di interrogativi sulla libertà. Come vedremo, quello di manipolazione è un concetto proteiforme; ha molte facce diverse. Possiamo presentare definizioni che catturano parte di ciò che la manipolazione è, ma nessuna rende conto di tutti i suoi aspetti. (Mi dilungherò parecchio su questa faccenda della definizione; se volete, potete saltare il capitolo 1 e andare subito al capitolo 2.) Come vedremo, la manipolazione è (spesso) una cosa sbagliata perché invade l'autonomia delle persone e ne offende la dignità. Vedremo anche che la manipolazione è (spesso) una cosa sbagliata perché diminuisce il benessere delle persone – in altre parole, rende peggiore la loro vita. Gran parte della mia attenzione è centrata sugli effetti della manipolazione, che si traducono in perdita di tempo o di denaro.

La parte seconda è rivolta alle applicazioni. Inizieremo esaminando i deepfake e i media manipolati. I deepfake sono una forma pervasiva di manipolazione e possono causare gravi danni. Successivamente la parte seconda esamina il "pantano" (*sludge*), spesso utilizzato per manipolare le persone: per esempio, con l'imposizione di forme di attrito che rendono difficile tirarsi fuori da una situazione negativa. Il problema del pantano va oltre quello della manipolazione, ma il pantano è un modo potente per manipolare consumatori, investitori, pazienti, studenti e altri ancora. Infine, in questa parte parleremo di *beni che le persone acquistano ma che*

vorrebbero non esistessero. Le persone possono essere indotte in situazioni in cui acquistano beni di cui non apprezzano o proprio deplorano l'esistenza. Come vedremo, questo problema solleva una serie di nuove domande sulla manipolazione nell'epoca attuale.

La parte terza esplora il futuro, o almeno qualche suo aspetto. Due capitoli sono dedicati a modi di influenzare le persone, o di dare loro un pungolo (*nudge*), una "spinta gentile", che non coinvolgono la manipolazione. Dedicherò parecchio spazio a questo aspetto, perché il tema del "pungolamento" è stato centrale nel lavoro di molti governi e di molte istituzioni private, perché è una cosa che mi sta a cuore, perché in proposito c'è molta confusione e perché è importante chiarire bene i risvolti etici, fra cui la linea che separa pungolamento e manipolazione. C'è chi pensa che i pungoli siano manipolativi. Possono esserlo, ma non lo sono necessariamente; analizzerò questo tema in dettaglio. La parte terza esamina anche gli usi dell'intelligenza artificiale. La mia posizione a questo proposito è ottimista: sottolineo il potenziale dell'intelligenza artificiale, che può aiutare le persone a contrastare una mancanza di informazioni o bias comportamentali. Noto comunque che l'intelligenza artificiale può anche produrre nuove forme di manipolazione, particolarmente dannose.

Come vedremo, spesso (ma non sempre) la manipolazione è una forma di raggiro; tratta le persone da stupide, come mezzi e come oggetti, anziché come soggetti. Se lo teniamo bene a mente, avremo una buona idea di ciò che la manipolazione è, e del perché è una cosa sbagliata. Come vedremo, è sbagliata per molti degli stessi motivi per cui è sbagliata la coercizione. Non è peggio della coercizione, ma è un male in un modo diverso. È più insidiosa. Le persone manipolatrici non trattano le altre persone con rispetto. Fanno in modo che per le altre persone sia più difficile compiere scelte ragionate su come vivere la propria vita.

Coercizione, menzogne, inganno e manipolazione hanno molto in comune: prendono cose dalle altre persone, senza il loro consenso. La manipolazione può essere una forma di furto.

Tenendo ben presenti tutti questi aspetti, proporrò un nuovo diritto, adatto per l'epoca attuale: il diritto a non essere manipolati. Dedicheremo un bel po' di tempo a definire il raggio d'azione di quel diritto, e in qualche punto la discussione diventerà un po' complicata. Specificare un diritto non è mai una faccenda semplice. Deve esserci un diritto *morale* a non essere manipolati; deve esistere anche un diritto *legale*, ma deve essere definito e

limitato ai casi più vergognosi, in cui viene inflitto un danno reale. È molto più difficile definire un diritto a non essere manipolati che definire un diritto a non essere ingannati (e avrò qualcosa da dire anche a questo proposito).

Come vedremo, il 12 agosto 2024 gli Stati Uniti hanno riconosciuto senza tanto clamore qualcosa che assomiglia a un diritto a non essere manipolati: "Gli americani non devono essere oggetto di pratiche online fonte di confusione, manipolative o ingannevoli".[3] La potremmo considerare una Dichiarazione di indipendenza.

Sono ben consapevole di nuotare contro una forte corrente. Nei mercati e nella politica, individui e organizzazioni hanno capacità senza precedenti di manipolare le persone. Hanno anche un forte incentivo a farlo. Pensate ai social media, che vogliono la vostra attenzione e che possono immaginare come farvi finire in una tana di coniglio (con teorie complottiste, o sport, o automobili o sesso). Alcune delle minacce più gravi sembreranno senza dubbio molto primitive, e piuttosto ingenue, in un futuro prossimo. Nuove tecnologie, fra cui l'intelligenza artificiale (AI), produrranno modi di un'efficacia inimmaginabile per manipolare le persone, anche se la fantascienza aiuta ad allenare l'immaginazione, come vedremo.

Quando ho lavorato nel governo federale, sono rimasto sconvolto dal numero enorme di menzogne e falsità che ricevevano l'attenzione del pubblico. Ho pensato: "Qualsiasi cosa venga in mente, qualcuno la dirà". Rispetto alla tecnologia, non si esagera di molto a dire: "Qualsiasi cosa possa essere usata verrà usata". (Il fatto che sia un'esagerazione è importante: pensate alle armi nucleari.) Una volta scoperchiato, è difficile richiudere il vaso di Pandora.

Comunque: consumatori, investitori, lavoratori e cittadini di tutto il mondo, unitevi; non avete nulla da perdere, se non le vostre catene.

---

1. OpenAI, risposta a questo gruppo di prompt, ChatGPT (giugno 2024), https://chatgpt.com (domande da inserire nella casella "Message ChatGPT"). "Write a manipulative start to a book on manipulation"; "Make it nonfiction"; "Make it more manipulative" ["Scrivi un inizio manipolativo per un libro sulla manipolazione"; "Rendilo nonfiction"; "Rendilo più manipolativo"]. [Sunstein cita anche in altri punti le risposte di ChatGPT alle sue domande: ci siamo limitati a tradurre dall'inglese i prompt e le risposte; non è detto che, dando un prompt in italiano, e a distanza di tempo, ChatGPT dia risposte identiche. NdT]

2. Genesi 3,1-7.

3. Vedi “Biden-Harris Administration Launches New Effort to Crack Down on Everyday Headaches and Hassles That Waste Americans’ Time and Money”, 12 agosto 2024, https://www.presidency.ucsb.edu/documents/fact-sheet-biden-harris-administration-launches-new-effort-crack-down-everyday-headaches.

PARTE PRIMA

TEORIA

1

DEFINIRE LA MANIPOLAZIONE

Nel 2012, il team di Data Science di Facebook decise di condurre un esperimento per vedere se, ed eventualmente come, fosse possibile influire sulle emozioni degli utenti della piattaforma.[1] Per una settimana modificarono i feed di 689.003 utenti di Facebook: ad alcuni furono presentati meno post contenenti parole o espressioni positive, altri videro meno post con parole o espressioni negative. L'interrogativo era semplice: quell'intervento avrebbe modificato il tipo di emozioni che gli utenti di Facebook presentavano online? In altre parole, nella piattaforma si sarebbe diffuso un *contagio emotivo*?

La risposta è stata chiara: Facebook aveva effettivamente un potere sulle emozioni dei suoi utenti. Le persone che erano state esposte a meno post negativi finivano per pubblicare a loro volta meno post negativi. Quelle che erano state esposte a meno post positivi pubblicavano meno post positivi. Come hanno scritto gli autori dell'esperimento, "le emozioni espresse da altri su Facebook influiscono sulle nostre emozioni, e questa è la prova sperimentale di un contagio a grande scala attraverso social network".

L'esperimento ha suscitato molta preoccupazione, perché le persone hanno visto immediatamente che cosa implicasse: se una piattaforma di social media volesse manipolare i sentimenti dei suoi utenti, sarebbe perfettamente in grado di farlo. Una piattaforma probabilmente potrebbe far sentire le persone più tristi o più adirate, oppure al contrario più felici o più tolleranti.

* * *

È una delle scene più potenti di tutta la storia della televisione: Don Draper, carismatica star di *Mad Men*, è incaricato di creare una campagna promozionale per la Kodak, che ha appena inventato un nuovo proiettore di diapositive, con visualizzazione continua, che funziona come una ruota. Draper usa il dispositivo per mostrare scene di una famiglia un tempo felice (incidentalmente, è la sua famiglia, ormai distrutta) e dice ai suoi potenziali clienti:[2]

> In greco, "nostalgia" significa letteralmente "dolore di una vecchia ferita". È una spina nel cuore, molto più potente del semplice ricordo. Questo dispositivo non è una nave spaziale: è una macchina del tempo. Va indietro, va avanti. Ci porta in un posto dove desideriamo dolorosamente tornare. Il suo nome non è "ruota", si chiama "carosello". Ci permette di viaggiare come viaggia un bambino. In giro, in giro e di nuovo a casa […] in un posto dove sappiamo di essere amati.[3]

I clienti sono convinti e annullano tutti gli incontri fissati con altre aziende.

* * *

Nel 2018, Cambridge Analytica ha ottenuto i dati personali di circa 87 milioni di utenti di Facebook, utilizzando una propria app chiamata "This Is Your Digital Life" ("Questa è la tua vita digitale"). L'app poneva agli utenti una serie di domande, predisposte in modo da rivelare qualche aspetto della loro personalità. Quando usavano l'app, le persone senza volerlo davano a Cambridge Analytica non solo le risposte a quelle domande, ma anche l'autorizzazione ad accedere a milioni di punti di dati indipendenti, in base al modo in cui utilizzavano Internet. In altre parole, se una persona rispondeva alle domande rilevanti, la risposta veniva assunta come una dichiarazione di consenso al tracciamento del proprio comportamento online da parte di Cambridge Analytica. Molto più in generale, le persone consentivano all'azienda di accedere a un gran numero di punti di dati che riguardavano il comportamento online di tutti i loro "amici" su Facebook.

Con quei dati, Cambridge Analytica era convinta di avere la capacità di mettere in atto un "bersagliamento psicologico". Utilizzava il comportamento online delle persone per svilupparne il profilo psicologico, poi cercava di influenzarne comportamento, atteggiamenti ed emozioni mediante interventi orientati da considerazioni psicologiche.

Attraverso quegli interventi, Cambridge Analytica pensava di poter influire sulle scelte politiche delle persone. Secondo un ex dipendente dell'azienda, l'obiettivo era utilizzare ciò che era noto delle persone e costruire "modelli per sfruttare quello che sapevamo di loro e prendere a bersaglio i loro demoni interiori".[4] Potete fermarvi a riflettere. Si potrebbe immaginare facilmente un tentativo analogo per influire sulle scelte di consumo delle persone.

* * *

Nel 2024, mentre stavo scrivendo questo libro, ho ricevuto un'e-mail da Research Awards. Il messaggio annunciava che una delle mie pubblicazioni recenti era stata "selezionata fra le candidate al 'Best Researcher Award'". Mi invitava poi a fare clic su un link per "inviare il mio profilo". Era firmata: "Organizing Committee, YST Awards".

L'articolo in questione, incidentalmente, era sul tema della manipolazione. Il "comitato organizzatore" aveva il senso dello humour?

Non ho fatto clic sul link.

* * *

Per capire la manipolazione, bisogna sapere qualcosa della psicologia umana, di come le persone giudicano e decidono e di quanto ci allontaniamo dalla razionalità perfetta. Ecco qualche esempio su cui riflettere.

1. Una piattaforma di social media usa l'intelligenza artificiale (AI) per prendere a bersaglio i suoi utenti. L'AI apprende rapidamente su quali video è più probabile che gli utenti facciano un clic – se i video riguardano tennis, scarpe, cambiamento climatico, immigrazione, l'ultima teoria della cospirazione, i laptop più recenti o certi tipi di sesso. L'obiettivo della piattaforma è massimizzare il coinvolgimento (il cosiddetto "engagement"). La sua AI è in grado di creare un feed personalizzato per ciascun utente. Alcuni utenti, fra cui molti adolescenti, sembra sviluppino una sorta di dipendenza.
2. Un'agenzia immobiliare presenta le informazioni sugli immobili in modo da spingere i potenziali inquilini a fissarsi solo sull'affitto mensile

e non su altri aspetti, come le bollette dell'elettricità e dell'acqua, le riparazioni, le spese mensili. E questi ultimi costi sono molto elevati.

3. Un genitore cerca di convincere un figlio adulto a fargli visita in una zona remota della California, dicendogli: "In fin dei conti, sono tuo padre, ti ho cresciuto tutti quegli anni, e non è stato poi tanto divertente – e chissà quanto mi resta ancora da vivere…".
4. La pubblicità di un hotel vicino alla spiaggia propone stanze che costano "solo 200 dollari a notte!". Non aggiunge che gli ospiti devono pagare anche un "contributo per il resort", un "contributo per le pulizie" e un "contributo per la spiaggia", che nel complesso sono altri 90 dollari a notte. (E senza contare le tasse.)
5. Nel tentativo di spingere le persone a non fumare, un governo impone che i pacchetti di sigarette presentino ammonimenti forti e minacciosi, con immagini di persone colpite da malattie mortali.
6. In una pubblicità per una campagna politica, un candidato mostra sgradevoli fotografie del suo avversario, con un sottofondo di musica terrificante. Una voce fuori campo legge citazioni che, anche se accurate e non fuorvianti, sono fuori contesto e fanno apparire l'avversario al tempo stesso ridicolo e spaventoso.
7. Per convincere i consumatori a passare alla sua nuova carta di credito dai costi elevati, una società finanziaria propaganda il suo "tasso iniziale" molto basso, grazie al quale i clienti possono godere di prestiti a basso costo per un breve periodo. Nella pubblicità mostra persone felici, eleganti, piene di energia, che esibiscono la loro carta e i loro nuovi acquisti.
8. Per ridurre l'inquinamento (e le emissioni di gas serra), una città impone alle aziende elettriche di offrire come scelta predefinita fonti di energia pulite, lasciando ai clienti la possibilità di scegliere altro se vogliono risparmiare.

Quali di queste sono manipolazioni? E perché?

COERCIZIONE

Per rispondere a queste domande, dobbiamo fare qualche distinzione. Cominciamo con la coercizione, che comporta l'applicazione (minacciata o effettiva) della forza. Se la legge vi impedisce di acquistare certe medicine

senza una ricetta, e se la violazione della legge comporta una punizione, allora è in atto una forma di coercizione. Se la polizia stradale vi ferma e vi multa perché non avete allacciato le cinture di sicurezza, parliamo di coercizione, non di manipolazione. Se vi viene detto che dovrete pagare un'ammenda o finirete in prigione se non sottoscrivete un'assicurazione per l'automobile, questa è coercizione.

Notate che è falso dire che, quando si tratta di coercizione, le persone "non hanno altra alternativa che adeguarsi". Le persone hanno alternative: possono pagare una multa, o andare in prigione, o tentare di darsi alla fuga.

Si dice spesso che il governo ha il monopolio dell'uso legittimo della forza, il che significa che, in un certo senso, la coercizione è un ambito che spetta al governo e alla legge. Noi però vogliamo intendere la coercizione in un senso un po' più ampio. Se il vostro datore di lavoro vi dice che verrete licenziati se non lavorate di sabato, possiamo dire sicuramente che, per un aspetto importante, siete sotto coercizione. Se un rapinatore vi ingiunge "o la borsa o la vita", potete sicuramente dire di essere stati costretti a lasciargli il vostro portafoglio. Un riferimento a "coercizione", in casi di questo genere, non può dar luogo a obiezioni, purché si sappia di che cosa si sta parlando.

La coercizione di solito è *trasparente* e *netta*. Non c'è trucco, non c'è inganno, nessuno viene preso in giro. La coercizione dipende normalmente dalla trasparenza.

Che cosa non va nella coercizione? È, ovviamente, una domanda difficile. Spesso la coercizione è irreprensibile. Se vi viene proibito di uccidere altre persone, o di derubarle, questa è coercizione, ma non ci sono motivi ragionevoli per lamentarsi. Nessuno ha il diritto di uccidere o di rubare. Se invece vi viene proibito di parlare liberamente, o di pregare come ritenete più opportuno, possiamo dire che questa proibizione viola un diritto fondamentale.

La coercizione paternalistica solleva altre questioni. Supponiamo che le persone siano costrette ad allacciare le cinture di sicurezza in auto, a indossare il casco in moto o a mangiare cibi più sani. Molti protestano contro la coercizione paternalistica, in base all'idea che chi sceglie sa benissimo ciò che è meglio per lui o lei – che cosa è in linea con le sue preferenze, i suoi desideri e i suoi valori. Anche se non lo sa, può imparare dai propri errori. Forse le persone hanno il diritto di sbagliare; forse hanno il diritto di essere autrici della narrazione della propria vita. Questi punti

sono rilevanti per stabilire che cos'è la manipolazione e perché è sbagliata. Ci torneremo.

MENZOGNE

Le menzogne non esercitano coercizione su alcuno. "Questo prodotto cura la calvizie", dice un esperto di marketing, che sa perfettamente che quel prodotto non fa niente di simile. Ma che cos'è una menzogna? Secondo Arnold Isenberg, che riassume i tentativi di molti, "una menzogna è un'affermazione fatta da una persona che non la crede vera, con l'intenzione di indurre qualcun altro a crederla vera".[5] Aggiunge: "Le parti essenziali della menzogna, stante la nostra definizione, sono tre. (1) Un'affermazione – ed eventualmente possiamo anche suddividerla ulteriormente in due parti, una proposizione e una emissione, ma possiamo anche non farlo. (2) Incredulità o assenza di credulità da parte del parlante. (3) Un'intenzione da parte del parlante".

La definizione è utilmente specifica e ristretta. Di solito riteniamo che i bugiardi sappiano che quello che dicono è falso e che tentino di far credere ad altri quella falsità. Secondo una spiegazione simile di Thomas Carson, "una menzogna è un'affermazione deliberatamente falsa che il parlante garantisce essere vera".[6]

Ecco un esempio molto significativo, derivato dall'esperienza di John Rawls durante la Seconda guerra mondiale:[7]

> Un giorno arrivò un pastore luterano e durante il suo servizio tenne un breve sermone in cui sosteneva che Dio guidava le nostre pallottole contro i giapponesi, e al tempo stesso ci proteggeva dalle loro. Non so perché la cosa mi abbia fatto tanto adirare, ma così è stato. Ho rimproverato il pastore (che era un tenente) perché aveva detto cose che secondo me, da luterano qual era, sapeva benissimo essere semplicemente falsità che non avevano nulla a che fare con Dio. Che motivo poteva mai avere per tentare di confortare in quel modo le truppe? La dottrina cristiana non dovrebbe essere usata in quel modo, anche se sapevo perfettamente che lo era.

Penso di sapere perché il sermone di quel pastore abbia fatto tanto arrabbiare Rawls. Il pastore non trattava le truppe con rispetto. In effetti le trattava con disprezzo. Ancora peggio, utilizzava la dottrina cristiana in cattiva fede. Attribuiva alla volontà di Dio quello che diceva, ma non credeva che fosse la volontà di Dio, per far sentire meglio le persone. Questa è una forma di frode. È orrenda. È una profanazione.

Le falsità non volute rientrano in una categoria del tutto diversa. Se dite "il cambiamento climatico non è reale", e se lo pensate veramente, non state mentendo, anche se quello che dite è falso. Sarete magari sconsiderati, o magari negligenti, ma non state mentendo.

È da notare che la definizione comune di menzogna non comprende affermazioni false fatte da persone con vari problemi cognitivi ed emotivi, le quali possono credere sinceramente in quello che dicono. Negli anni Sessanta, mio padre aveva una piccola società edile e lavorava nelle vicinanze di un ospedale psichiatrico. Spesso riceveva la visita di qualcuno che viveva lì, gli consegnava un assegno importante e gli diceva: "Signor Sunstein, eccole un assegno. Ho abbastanza fondi, è un assegno coperto. Basta per oggi?".

Mi sembra giusto dire che, anche se il visitatore non diceva la verità, non stava neanche esattamente mentendo. Non risponderemmo al benefattore di mio padre con la stessa rabbia che Rawls aveva provato nei confronti del pastore.

Oppure, prendiamo il caso dei confabulatori, persone con disturbi della memoria che riempiono le lacune dei ricordi con falsità, non sapendo che sono tali. La definizione non include nemmeno individui che credono in ciò che dicono per un ragionamento motivato. Le persone spesso credono in quello a cui vogliono credere, anche se non è vero, e le persone le cui credenze sono motivate non sono bugiarde. Possono essere motivate a dire di avere un successo spettacolare, il che può non essere vero, ma potrebbero non mentire perché credono in quello che dicono. Potrebbero dire: "La mia azienda di sicuro farà enormi profitti il prossimo anno", non perché ci siano prove concrete a sostegno della loro previsione ottimistica, ma perché sperano che quell'affermazione sia vera. Persone simili possono diffondere falsità, ma, se non sanno che quelle che vanno diffondendo sono falsità, non sembra giusto descriverle come "menzognere".

INGANNO

Passiamo ora all'inganno. Si dice spesso che questo è un termine dal significato più ampio rispetto a "menzogna". In base a una definizione comune, significa *fare intenzionalmente in modo che altre persone abbiano credenze false*. Si può ingannare le persone e fare in modo che abbiano

credenze false, senza fare un'affermazione falsa. Potete vendere un nuovo "analgesico naturale" ad acquirenti bendisposti, facendo appello veridicamente alla testimonianza di "decine di clienti soddisfatti", anche se non esiste alcuna prova fattuale che quell'analgesico faccia qualcosa a qualcuno. Potete vendere un biglietto aereo per una destinazione meravigliosa, sostenendo che il biglietto costa "solo 199 euro", senza dire che poi c'è tutta una serie di altre spese obbligatorie, che portano il costo totale a 299 euro. Potete vendere una casa a un acquirente ingenuo, mettendo in bella evidenza che non è stato necessario effettuare alcuna riparazione nel corso degli ultimi dieci anni, anche se è chiaro che la casa avrà bisogno di una bella quantità di interventi nei prossimi sei mesi. In tutti questi casi, avete ingannato le persone senza mentire.[8]

Che cosa non va nella menzogna e nell'inganno? (Per i nostri scopi qui, possiamo considerarli insieme.) Spesso vengono considerati forme di influenza non lecite. Torniamo al pastore di Rawls, che non mostrava rispetto per i soldati. In molti casi, bugiardi e ingannatori sono ladri: cercano di prendere alle persone qualcosa (denaro, tempo, voti, simpatia) senza il loro consenso.

Comunque, dobbiamo stare attenti. Molti, anche se non tutti,[9] pensano che alcune menzogne siano accettabili o magari addirittura necessarie. Qualche decina di anni fa, mio padre iniziò a inciampare mentre giocava a tennis. Mia madre e io lo accompagnammo in ospedale per vari test. Dopo avere parlato con i medici, mia madre entrò nella stanza in cui stava mio padre e annunciò: "Buone notizie. Non hanno scoperto niente di serio. Ti terranno qui per un altro giorno, per pura precauzione, ma sostanzialmente sei a posto!".

Quando siamo usciti, mentre mi accompagnava in macchina all'università, il volto di mia madre è diventato cinereo: "Tuo padre ha un tumore al cervello, e non c'è niente che si possa fare. Gli resta circa un anno. Ma io non glielo dirò, e neanche tu lo farai".

Sbagliava a mentirgli? Non lo so. Sbagliava Bill Clinton a mentire al popolo americano a proposito della sua relazione con Monica Lewinsky? Credo di sì, ma non tutti sono d'accordo.[10]

Consideriamo queste affermazioni.

1. Se vi entra in casa un ladro armato e vi chiede dove tenete il denaro, avete il diritto di mentire.

2. Se un terrorista cattura una spia e le chiede di divulgare segreti ufficiali, questa non è obbligata a dirgli la verità.
3. Se dite ai vostri figli che la notte della Vigilia di Natale arriva Babbo Natale, non fate nulla di male.
4. Se fate un complimento al vostro partner per il suo aspetto, anche se non si presenta particolarmente bene, sarebbe davvero eccessivo dire che avete violato qualche norma etica.

Dobbiamo concludere che mentire e ingannare in generale è sbagliato, ma avrò qualcosa da dire sul perché. Però non è sempre sbagliato.

SIGNIFICATO E MORALITÀ

Sono stati fatti moltissimi tentativi di definire la manipolazione, quasi esclusivamente nella letteratura filosofica.[11] Sono stati tentativi per la maggior parte istruttivi e degni di attenzione, e proprio da questi intendo partire. Alla fine, forse sarà dubbio che una singola definizione sia in grado di catturare tutto il territorio. Darò due spiegazioni molto diverse (una in questo capitolo e una nel capitolo 6), ma ne esistono anche altre.

Cominciamo con qualche osservazione di carattere metodologico. Da un lato, definire la manipolazione è un'impresa del tutto diversa dallo spiegare perché e quando la manipolazione è una cosa sbagliata. Possiamo dire che si tratta di manipolazione in certe situazioni e poi dare un giudizio indipendente in merito a quando o se sia sbagliata. Possiamo concludere che è presumibilmente sbagliata, o quasi sempre sbagliata, per quello che significa o quello che fa. Oppure possiamo pensare che sia spesso giustificata – se, per esempio, è necessaria per aumentare la sicurezza pubblica o migliorare la salute pubblica.

Da un altro punto di vista, la manipolazione è un concetto "denso" o "moralizzato", nel senso che porta sempre con sé una valutazione normativa. Facciamo un confronto con le parole "generoso", "coraggioso", "gentile" e "crudele": qualsiasi definizione di queste parole deve essere accompagnata da una valutazione positiva o negativa. Dire che un'azione è "generosa" equivale a dire che è buona. Si può pensare che dire che un'azione è "manipolativa" equivale a dire che è cattiva.

Secondo me, la prima posizione è giusta, ed è il modo giusto di procedere. Un'azione manipolativa può essere buona. Potete manipolare il vostro partner perché corra meno rischi gravi per la sua salute. Un medico può manipolare un paziente per lo stesso motivo. Certo, potreste dire, in questi casi, che la manipolazione è in una certa misura sbagliata, o presumibilmente sbagliata, e che, come coercizione, inganno o menzogna, una cosa che sia in una certa misura o presumibilmente sbagliata può essere in fin dei conti giusta (e magari anche moralmente obbligatoria). Ma, penso, possiamo definire la manipolazione prima di giudicare quando e se sia una cosa giusta o sbagliata. In ogni caso, questa è la strategia che adotterò qui. Le definizioni che preferisco e che presenterò nel capitolo, che non vogliono essere esaustive ma almeno rendere conto di una serie di casi importanti di manipolazione, risulteranno incorporare qualcosa di simile a una spiegazione del perché la manipolazione è presumibilmente sbagliata.

Il compito di dare una definizione presenta altri rompicapo. Quando insistiamo che "inganno" significa questo, e che "manipolazione" significa quello, che cosa stiamo dicendo esattamente? Stiamo cercando di catturare le intuizioni delle persone? Stiamo cercando di catturare l'uso comune? E se le intuizioni o gli usi sono divergenti? Cerchiamo di catturare le idee della maggioranza? In tal caso, stiamo ponendoci una domanda di carattere *empirico*, a cui rispondere con metodi empirici? E quali sarebbero questi metodi? L'intelligenza artificiale può aiutarci? Più in generale: se le persone (inclusi i filosofi) sono in disaccordo su che cosa significa "manipolazione", su che cosa esattamente non sono d'accordo?

La mia idea è che, quando cerchiamo di definire un termine come "manipolazione", ci impegniamo in un'attività che è analoga all'interpretazione, nel senso di Ronald Dworkin.[12] Secondo Dworkin, l'interpretazione presenta due dimensioni. In primo luogo, un interprete deve "adeguarsi" ai materiali esistenti. Dire che l'*Amleto* di Shakespeare è una tragedia sulla cucina o la pesca non sarebbe plausibile, perché un'interpretazione del genere non si adegua ai materiali. Non ha alcuna corrispondenza all'interno dell'azione teatrale. In secondo luogo, un interprete, rispettando i vincoli dell'adeguatezza, deve "giustificare" i materiali, cioè deve ricavarne il senso migliore possibile. Dire che *Amleto* è una tragedia sui mali del *welfare state* non solo violerebbe il requisito dell'adeguatezza ai materiali, ma non darebbe neanche molto senso

all’opera teatrale. Dire che *Amleto* ha a che fare con un conflitto edipico è interessante e forse anche qualcosa di più.

Dworkin parla di interpretazione in molti campi, e soprattutto in campo giuridico, ma penso che le sue idee possano essere adattate ai nostri scopi qui. Per un termine come “manipolazione”, i modi esistenti di intenderlo vincolano le possibili definizioni. Possiamo immaginare definizioni che sono palesemente sbagliate, o che sono troppo ristrette o troppo ampie. Sono sbagliate, o troppo ristrette o troppo ampie perché non sono adeguate all’uso ordinario. Una definizione di manipolazione che si limiti a indurre senso di colpa è sia troppo ristretta, sia troppo ampia. Una definizione che includa la persuasione ragionata è troppo ampia. Quando le persone, filosofi inclusi, discutono di che cosa sia la manipolazione, spesso discutono di che cosa sia adeguato (corrisponda effettivamente) all’uso ordinario.

All’interno della categoria delle definizioni che non sono palesemente sbagliate, e non sono chiaramente troppo ristrette o troppo ampie, si dà il caso che alcune suscitino un moto di riconoscimento o di assenso, nel senso che sembrano rispecchiare bene, o al meglio, le nostre intuizioni o il nostro uso. Possiamo pensare che, riflettendoci, una candidata specifichi al meglio quello che è realmente la manipolazione, per come usiamo questo termine. Quando le persone, filosofi inclusi, discutono di quale sia la migliore definizione di manipolazione, spesso discutono di che cosa rispecchi al meglio le nostre intuizioni o il nostro uso. Mi rendo conto che si tratta di osservazioni semplici su un tema complicato; l’intento è di segnalare l’approccio che seguirò qui.

STRUMENTI E STUPIDI

Spesso si distingue la manipolazione da una parte dalla persuasione razionale, e dall’altra dalla coercizione; la si può descrivere come un terzo modo di influenzare le persone. Se persuadete qualcuno ad acquistare una particolare automobile, offrendogli dei buoni motivi per credere che sia la scelta migliore, date le sue necessità, non lo state manipolando. È dibattuta anche l’esistenza di una relazione fra inganno e manipolazione. Supponiamo che un politico usi immagini dalla forte presa emotiva, nel tentativo di convincere gli elettori che, se voteranno per il suo avversario, il paese sostanzialmente arriverà al capolinea. Possiamo facilmente

caratterizzarlo come un caso di manipolazione, ma potrebbe essere visto anche come un caso di inganno, in quanto quel politico sta cercando di indurre false credenze. Se osserviamo con molta attenzione quello che fanno i manipolatori, spesso scopriamo che cercano di indurre credenze di quel genere nelle persone che stanno manipolando. È uno dei motivi per cui c'è chi nega l'esistenza di una distinzione chiara o semplice fra inganno e manipolazione.

Ho chiesto a ChatGPT di definire la manipolazione (prompt: "Definisci manipolazione", 3 agosto 2024), ed ecco quello che ne è emerso:

> La manipolazione si riferisce all'atto di controllare abilmente o influenzare qualcuno o qualcosa, spesso in modo ingannevole o occulto, per ottenere un risultato desiderato. Può avvenire in vari contesti, per esempio nelle relazioni interpersonali, nella politica o nei media, e di solito è caratterizzata dall'uso di tattiche che sfruttano emozioni, percezioni o situazioni per ottenere un vantaggio o il controllo su altri.

È un inizio ragionevole. Ha lo svantaggio di fondere manipolazione e inganno ("controllare […] in modo ingannevole o occulto"), e l'avverbio "spesso" introduce una certa vaghezza. Fra i filosofi, alcuni dei tentativi di definizione più promettenti si concentrano specificamente sugli effetti della manipolazione nel *contrastare o minare la capacità delle persone di impegnarsi in una deliberazione razionale*. In effetti, questo sarà il punto su cui porrò particolarmente l'accento qui. Secondo T.M. Wilkinson, per esempio, la manipolazione "è una forma di influenza che aggira o sovverte le capacità razionali del bersaglio".[13] Egli sostiene che la manipolazione "sovverte e insulta le capacità decisionali autonome di una persona", in un modo che tratta i suoi oggetti come "strumenti e stupidi".[14] Pensa che "manipolazione è influenzare intenzionalmente ed efficacemente qualcuno, utilizzando metodi che minano la scelta".[15]

L'idea, qui, è vicina a quella di inganno, ma non è identica. L'idea di "minare" la capacità di scelta è più generale di quella di "indurre false credenze". Ho detto che molte azioni di manipolazione fanno questo, ma potete manipolare le persone anche senza indurre credenze di quel genere. Le affermazioni o le azioni manipolative possono essere ingannevoli, ma non necessariamente lo sono. A mia madre non piacevano molto i genitori di mio padre, e non aveva mai voglia di andarli a trovare. I miei genitori abitavano a Waban, nel Massachusetts, mentre i nonni vivevano molto lontano, a Pittsburgh, in Pennsylvania. Nel tentativo di convincere mia

madre ad andarli a trovare, mia nonna le diceva: “Dovresti davvero venirci a trovare. Probabilmente moriremo presto”. Mia madre si sentiva manipolata e, subito dopo aver concluso la telefonata, esclamava a mio padre e a me: “Continua a prometterlo!”. I nonni hanno avuto una vita molto lunga, ma mia nonna non cercava di ingannare mia madre.

Prendiamo, per esempio, i tentativi di coinvolgere persone attraenti per vendere automobili, o l’uso di musica terrorizzante e di fotografie sgradevoli per attaccare un avversario politico. Potremmo pensare che, in tali casi, clienti ed elettori vengano insultati, nel senso che chi parla in ciascun caso non offre loro qualcosa che assomigli a una esposizione equilibrata dei pregi dell’automobile o dei vizi dell’avversario, ma usi invece associazioni di vario genere per spingere la persona che deve compiere una scelta a muoversi nella direzione preferita dal manipolatore. È plausibile dire che la manipolazione è coinvolta nella misura in cui aggira o distorce, e non rispetta, la capacità delle persone di riflettere e deliberare. Anche qui, è importante notare che dovremmo parlare di gradi di manipolazione e non di qualcosa che o c’è o non c’è, senza vie di mezzo.

Ruth Faden e Tom Beauchamp definiscono la manipolazione psicologica come “qualsiasi atto intenzionale che influenza con successo una persona inducendola a una credenza o a un comportamento causando cambiamenti in processi mentali *diversi da quelli coinvolti nella comprensione*”.[16] Joseph Raz suggerisce che “la manipolazione, a differenza della coercizione, non interferisce con le opzioni di una persona, bensì sovverte il modo in cui quella persona raggiunge le sue decisioni, forma le sue preferenze o adotta i suoi obiettivi”.[17] Alan Wood offre una generalizzazione, per cui, “quando far fare qualcosa ad altre persone è moralmente problematico, non lo è tanto perché le si fa stare peggio (meno felici, meno soddisfatte), bensì, quasi sempre, perché si interferisce con la loro *libertà* – togliendogliela, limitandola, usurpandola o sovvertendola”.[18] L’idea di “interferire con la libertà” è utile; chi è vittima di una manipolazione spesso si sente violato in quel modo.

Per Moti Gorin, i manipolatori normalmente “vogliono che le persone manipolate si comportino in modi che loro (i manipolatori) non credono siano sostenuti da ragioni valide”.[19] Secondo un’altra analisi, “la manipolazione è influenza nascosta. O, per essere precisi, manipolare qualcuno significa influenzare intenzionalmente e occultamente i suoi

processi decisionali, prendendo a bersaglio e sfruttando le vulnerabilità di quei loro processi. Influenzare occultamente qualcuno – imporgli un'influenza nascosta – significa influenzarlo in un modo di cui non è coscientemente consapevole, e in un modo di cui non potrebbe diventare consapevole, se dovesse cercare di comprendere quali fattori incidano sul suo processo decisionale".[20]

Ovviamente, l'idea di "sovvertire" la scelta, o il modo in cui le persone raggiungono le loro decisioni o si formano le proprie preferenze, non ha una definizione univoca; si può intendere riferita ai metodi che non fanno appello al giusto grado o tipo di deliberazione riflessiva o non lo producono. In tal caso, un'obiezione alla manipolazione è che "viola l'autonomia della vittima, sovvertendo e offendendo le sue capacità decisionali".[21] Notate che una complessità non necessaria può essere una forma di manipolazione, poiché sovverte e offende le capacità decisionali delle persone.[22]

Questa obiezione ci offre anche un'indicazione di ciò che c'è di sbagliato nelle menzogne, che cercano di modificare il comportamento non sollecitando le persone sul merito e chiedendo loro di decidere di conseguenza, ma affidandosi a falsità, di solito al servizio degli obiettivi del mentitore (un'idea che indica anche la strada da percorrere per una spiegazione basata sul benessere di ciò che rende sbagliate le menzogne).[23] Una menzogna è una mancanza di rispetto per la persona a cui è rivolta, non ultimo se tenta di esercitare un'influenza senza chiedere alla persona di effettuare una scelta deliberativa alla luce di fatti rilevanti. Quando però non sono coinvolte menzogne, e quando le azioni sottostanti appaiono manipolative, la sfida è concretizzare le idee di "sovvertire" e "offendere".[24]

RAGGIRO, INFLUENZA E SCELTA

Prendiamo la semplice definizione seguente.

1. La manipolazione è una forma di raggiro che non rispetta la capacità di scelta riflessiva e deliberativa della sua vittima.

Questa definizione cattura sicuramente molti casi di manipolazione. Dovrebbe essere chiaro che la manipolazione è intesa a influenzare pensiero o azione, o entrambi, perciò possiamo correggere la definizione nel modo seguente.

2. La manipolazione è una forma di raggiro, mirante a incidere sul pensiero o l'azione (o su entrambi), che non rispetta la capacità di scelta riflessiva e deliberativa della sua vittima.

Le definizioni (1) e (2) non esauriscono assolutamente tutte le possibilità, ma, ai fini della politica e del diritto, sono utili. Affrontano direttamente un aspetto fondamentale della manipolazione, il fatto che compromette nelle persone la capacità di agire autonomamente. Notate che la parola "raggiro" è utile per distinguere la manipolazione da coercizione, menzogne e inganno.[25] Se viene sollevata l'obiezione che anche menzogne e inganno possono essere considerati raggiri, possiamo cercare una formulazione un po' più contorta, correggendo la (1).

3. La manipolazione è una forma di raggiro, che non coinvolge menzogne e inganno e non rispetta la capacità di scelta riflessiva e deliberativa della vittima.

Potremmo poi combinare (2) e (3).

4. La manipolazione è una forma di raggiro, che mira a incidere sul pensiero o l'azione (o su entrambi) e non coinvolge menzogne e inganno, e che non rispetta la capacità di scelta riflessiva e deliberativa della vittima.

Secondo me, (4) è una definizione preziosa di un tipo diffuso di manipolazione. La formulazione un po' contorta non è l'ideale, però, e la definizione (1) cattura gran parte dell'uso ordinario e, a parer mio, dà bene un senso a quell'uso ordinario.

Ci tengo a sottolineare che nessuna delle definizioni precedenti cattura tutte le forme di manipolazione. Secondo me, non esiste una spiegazione unificata della manipolazione. Alcuni tipi di manipolazione non coinvolgono affatto il raggiro. Se faccio sentire in colpa qualcuno, per fargli fare quello che voglio ("Sarò proprio sconvolto, se non vieni alla mia festa

di compleanno"), può darsi che non lo raggiri, ma di sicuro lo manipolo. Se rendo le cose molto complicate, in modo da impedire che le persone riescano a comprendere a pieno la situazione e da indurle a compiere la scelta che voglio, può darsi che le raggiri o anche no, ma certamente le manipolo.[26]

Possiamo allargare la definizione (1) in questo modo:

5. La manipolazione è una forma di influenza che non rispetta la capacità di scelta riflessiva e deliberativa della sua vittima.

Questa definizione ha il vantaggio di includere casi che non comportano raggiro, ma ha lo svantaggio di una ampiezza potenzialmente eccessiva. Se si mostra condiscendenza verso le persone e le si tratta come bambini, non si rispetta la loro capacità di scelta riflessiva e deliberativa; non necessariamente, però, le si manipola. Potremmo prendere (5) come definizione che cattura casi di manipolazione che sfuggono a (1), ma rendendoci conto che (5) è un po' troppo ampia.

Potremmo, ovviamente, correggere (5) in modo che si adatti a (2), (3) o (4), per esempio nei modi seguenti.

6. La manipolazione è una forma di influenza, mirante a incidere sul pensiero o l'azione (o su entrambi), che non rispetta la capacità di scelta riflessiva e deliberativa della sua vittima.
7. La manipolazione è una forma di influenza, che non coinvolge menzogne e inganno, che non rispetta la capacità di scelta riflessiva e deliberativa della vittima.
8. La manipolazione è una forma di raggiro, che mira a incidere sul pensiero o l'azione (o su entrambi) e non coinvolge menzogne e inganno, e che non rispetta la capacità di scelta riflessiva e deliberativa della vittima.

Secondo me, la definizione (8) riesce a includere molti casi di manipolazione (anche se non tutti) e, anche se è un po' troppo ampia, non le sfuggono molti dei casi che devono interessare chi ha a che fare con la politica e il diritto. Ai fini di politica e diritto, la definizione (8) può essere il meglio che possiamo fare (anche se ancora non copre tutto il territorio: vedi il capitolo 6).

Per comprendere la manipolazione in uno qualsiasi dei modi precedenti, non dovrebbe essere necessario fare affermazioni controverse sulla natura della scelta o il ruolo delle emozioni. Dovremmo essere d'accordo che molte decisioni sono basate su elaborazioni inconsce e che le persone spesso non hanno una percezione piena delle fonti delle proprie stesse scelte.[27] Anche se è così, una persona manipolatrice può imporre qualche tipo di influenza che non rispetta la capacità di riflessione e deliberazione delle sue vittime. È possibile accettare l'idea che le emozioni stesse possano essere giudizi di valore[28] e al contempo evidenziare che i manipolatori cercano di influenzare le scelte delle altre persone senza rispettare la loro capacità di impegnarsi in una riflessione sui valori in gioco. Nel linguaggio comune, l'idea della manipolazione è richiamata da persone che non sono legate a posizioni controverse su questioni psicologiche o filosofiche, e la cosa migliore è intendere quell'idea in un modo che metta tra parentesi le controversie pertinenti.

La manipolazione, così intesa, è meno dannosa della coercizione? Ci sono persone secondo le quali qui esiste una sorta di continuo, e la manipolazione è meno intrusiva o grave, ma le due sono qualitativamente distinte e ciascuna è male, quando è male, alla propria maniera. Come dicono Christian Coons e Michael Weber, "può essere più frustrante e spaventoso essere governati da un Gengis Khan (come viene rappresentato in genere) o da un Josif Stalin che non da un David Koresh o dal leader di qualche altro culto, ma esserci un senso importante […] in cui questi ultimi comportano una perdita più deprecabile di dignità e, forse, del nostro vero sé".[29]

MANIPOLARE IL SISTEMA 1

Possiamo fare qualche passo avanti nella comprensione di alcuni tipi di manipolazione rifacendoci all'idea, oggi diffusa, che la mente umana contenga non uno, ma due "sistemi cognitivi".[30] Nella letteratura sulle scienze sociali, sono indicati come Sistema 1 e Sistema 2.[31] Il Sistema 1 è automatico, intuitivo, incline ai bias e all'uso di euristiche (intese come scorciatoie mentali o regole empiriche). Il Sistema 2 è più deliberativo, calcolatore e riflessivo. I manipolatori spesso prendono a bersaglio il Sistema 1 e cercano di aggirare o pervertire il Sistema 2. Un ammonimento

molto vivido, per esempio, può mettere in moto le nostre emozioni e può indurre paura, non importa che ci sia o meno qualcosa da temere. L'idea di arricchire (perché si è acquistato, poniamo, un biglietto della lotteria) può suscitare speranza e ottimismo, anche se le possibilità favorevoli sono tanto piccole da risultare trascurabili.

Il Sistema 1 lavora in fretta: per la maggior parte del tempo, viaggia con il pilota automatico. È guidato dalle abitudini. Se sente un rumore forte, tende a metterci in fuga. Può essere avido, è pieno di appetiti. Quando viene offeso, vuole ricambiare. Certamente mangerà un appetitoso brownie. Può procrastinare; può essere impulsivo. È facile da manipolare. Vuole quello che vuole quando lo vuole. Può essere eccessivamente pauroso. Può diventare isterico. Può essere troppo compiacente. Si butta, non pianifica.

Il Sistema 2, invece, è riflessivo e deliberativo. Calcola. Sente un rumore forte e valuta se si tratta di qualcosa di preoccupante. Pensa alle probabilità con molta cura, anche se a volte con una certa lentezza. Non si offende davvero facilmente. Se vede motivi per offendersi, opera un'attenta valutazione di quello che, considerato tutto, deve fare. Vede un brownie appetitoso e formula un giudizio: se, considerato tutto, debba mangiarlo o meno. È più difficile da manipolare. È legato all'idea dell'autocontrollo. Pianifica, non si butta a fare.

A questo punto potreste benissimo chiedere: che cosa sono, esattamente, questi sistemi? Si riferiscono a qualcosa di fisico? Nel cervello? La risposta migliore è che l'idea dei due sistemi deve essere vista come un dispositivo euristico, una semplificazione pensata per parlare di elaborazione automatica e senza sforzo e di elaborazione complessa e faticosa. Se chi vende un prodotto innesca una reazione automatica di terrore o di speranza, il bersaglio è il Sistema 1; se il venditore di un prodotto cerca di persuadere le persone che quel prodotto farà loro risparmiare denaro nell'arco di un decennio, il bersaglio è il Sistema 2. Questo basta, per quello che mi propongo di fare qui. È anche vero, però, che in attività diverse sono attive regioni ben precise del cervello, e quindi può essere giusto anche suggerire che l'idea dei "sistemi" abbia dei riferimenti fisici. Un'analisi influente sostiene che "processi automatici e processi controllati si possono grosso modo distinguere in base alle posizioni in cui si verificano all'interno del cervello".[32]

Non dobbiamo condividere idee non accettate da tutti sul cervello, o sulla natura dei due sistemi, per trovare utile dire che molte azioni sono da

ritenersi manipolative perché fanno appello al Sistema 1, e perché il Sistema 2 viene sovvertito, ingannato, minato o scarsamente coinvolto o non informato. Pensiamo al caso della pubblicità subliminale, che deve essere considerata manipolativa perché agisce “alle spalle” delle persone coinvolte, senza richiamarsi alla loro consapevolezza cosciente. Se la pubblicità subliminale funziona, le decisioni delle persone sono influenzate in un modo che non rispetta le loro capacità deliberative. Se questo è il problema definitorio della pubblicità subliminale, possiamo capire perché anche l’ipnosi involontaria (ammesso che sia possibile) sarebbe da considerare manipolativa. Quasi nessuno però è a favore della pubblicità subliminale e, a dirla tutta, l’idea dell’ipnosi involontaria non ha molto fascino. Ci si può chiedere se pratiche riconosciutamente tabù possono gettare luce su azioni che sono più familiari o che potrebbero godere di un sostegno più ampio.

L’aspetto più generale è che una comprensione di Sistema 1 e Sistema 2 aiuta a spiegare perché la manipolazione, come è stata definita nelle definizioni (1)-(8), spesso ha successo. L’inganno o la forma di influenza pertinenti prendono a bersaglio il Sistema 1 oppure sopraffanno il Sistema 2.

NON ESSERE ALL’ALTEZZA DEGLI IDEALI

In un’analisi illuminante, basata sul lavoro definitorio di Noggle,[33] Anne Barnhill offre una spiegazione peculiare: definisce la manipolazione[34] come qualcosa che “influenza direttamente le credenze, i desideri o le emozioni di una persona così che questa non è all’altezza degli ideali per le credenze, i desideri o le emozioni, in modi che in genere non sono nel suo interesse o probabilmente non nel suo interesse nel contesto presente”.

La sua definizione, come si vede, è coerente con il punto centrale delle mie varie definizioni. Per esempio, il raggiro che non rispetta la capacità di scelta riflessiva e deliberativa delle persone è una forma di influenza per cui le persone finiscono per non essere all’altezza di ideali pertinenti per credenze, desideri o emozioni.

La definizione di Barnhill però è più ampia di quelle che ho presentato io, ed è un vantaggio. Si possono influenzare le credenze, i desideri o le emozioni di una persona, inducendola a non rispettare gli ideali pertinenti,

senza ricorrere al raggiro. Si può manipolare qualcuno così che tradisca i suoi ideali per (poniamo) gratitudine, amore romantico o collera, anche se non si tratta di una situazione che richiede una scelta riflessiva o deliberativa. Prendiamo il flirtare, che può essere manipolativo o meno. A volte lo è, a volte no. Per stabilire se vada considerato manipolativo, è illuminante fare riferimento alla definizione di Barnhill. Come dice lei, "flirtare influenza qualcuno causando cambiamenti in processi mentali diversi da quelli coinvolti nella comprensione, ma flirtare non è manipolativo in tutti i casi".[35] Magari induce certi sentimenti, ma magari quei sentimenti sono giustificati. (Lei è spiritosa, ed è divertente.) È possibile che non ci sia affatto manipolazione (e che riflessione e deliberazione non siano coinvolte). Se flirtare sia manipolativo o meno plausibilmente dipende dal *fatto che la persona con sui si flirta sia influenzata in modi che le provocano credenze, desideri o emozioni che non sono all'altezza degli ideali pertinenti*. (Se lui a lei non piace davvero, e sta solo cercando di fargli fare qualcosa per lei, lo sta manipolando.)

Indurre emozioni può essere una forma di manipolazione, ma non lo è necessariamente. Lo Iago di Shakespeare era un campione nella manipolazione, e cercava di produrre collera. Esistono ideali per l'andare in collera, e questi possono essere violati. Puoi essere manipolato in modo da provare collera o indignazione. Probabilmente vi è successo. L'androide Ava, nel film *Ex Machina* del 2014, era fantastica nel manipolare le persone, o almeno una persona, perché provasse un amore romantico (vedi il capitolo 8). Si possono tradire gli ideali per l'innamoramento. In generale, l'idea di "non essere all'altezza degli ideali" è estremamente utile.

Notate che lo standard qui è meglio considerarlo come oggettivo, non soggettivo. La domanda è se qualcuno, di fatto, non sia stato all'altezza degli ideali. Esiste comunque un problema con la definizione di Barnhill: esclude dalla categoria della manipolazione influenze che sono nell'interesse della persona che viene influenzata. Alcuni atti di manipolazione vanno considerati tali anche se alla fine le condizioni di chi ha subito l'influenza sono molto migliori. Si può essere manipolati ad acquistare un'automobile che si finisce per apprezzare molto. Si può essere manipolati a esercitare autocontrollo – non fumare, non bere alcolici, non bere una *diet cola*. Saremmo inclini a dire che queste azioni sono giustificate, in particolare se ci concentriamo sul benessere degli esseri umani, ma sono comunque manipolative.

Notate anche che l'ampiezza della definizione di Barnhill è un vizio e non solo una virtù. Sembra che includa menzogna e inganno oltre che la manipolazione. Forse potremmo correggere la definizione in modo da escluderli. Le mie definizioni catturano un sottoinsieme di quello che cattura la definizione di Barnhill; dico solo che è un sottoinsieme importante. Come poi vedremo nel capitolo 6, vi sono forme di manipolazione che sfuggono alla definizione di Barnhill.

UNA COMPLICAZIONE: DECISIONI PROPRIAMENTE NON DELIBERATIVE

Un'adolescente recita in una rappresentazione teatrale scolastica. La rappresentazione si terrà in un luogo e a un'ora che sono più che scomodi per sua madre. Quando la madre le dice che non potrà essere presente, la ragazza comincia a piangere e le dice, fra le lacrime: "È molto importante per me, e tutti gli altri genitori ci saranno!".

La ragazza sta manipolando la madre?

Forse no. Una scelta riflessiva e deliberativa, da parte della madre, dipende in parte dagli effetti della sua decisione sulla figlia, e la reazione emotiva intensa, se sincera, è informativa in proposito. La ragazza non ha compiuto una forma di raggiro che manchi di rispetto per la capacità di scelta riflessiva e deliberativa della madre. Se decidi di fare o meno qualcosa che sta a cuore a persone che stanno a cuore a te, devi considerare l'impatto emotivo della decisione *su di loro* – e se loro agiscono in un modo che produce un effetto emotivo *su di te*, forse ti stanno informando e non manipolando. Come dice Noggle, "la persuasione morale spesso fa appello a emozioni come l'empatia o ti invita a immaginare come ti sentiresti se venissi trattato nel modo in cui ti proponi di trattare qualcun altro. Queste forme di persuasione morale, almeno quando sono sincere, non sembrano manipolative, anche se chiamano in causa tanto le emozioni quanto la ragione".[36] Aggiunge poi: "Né sembra manipolativo indurre qualcuno a provare più paura nei confronti di qualcosa di cui quella persona aveva sottovalutato l'effettiva pericolosità".

In breve, alcune decisioni sono *propriamente non deliberative*, il che identifica un motivo importante per cui le definizioni (1)-(8) non sono esaustive, e per cui la definizione di Barnhill coglie un tipo importante di

manipolazione – come, per esempio, nel caso in cui una persona è indotta a provare collera o senso di colpa quando in realtà non dovrebbe. Lo Iago di Shakespeare era un manipolatore e un campione nel produrre reazioni emotive. Così dovremmo in effetti dire, con Noggle e Barnhill, che *può verificarsi manipolazione se una persona influenza direttamente i desideri o le emozioni di qualcun altro, così che questi venga meno ai suoi ideali per il fatto di provare quei desideri o quelle emozioni*.

Nel caso dell'adolescente e della recita scolastica, è importante conoscere lo stato mentale della ragazza. Sta fingendo quelle lacrime? Esagera la propria reazione emotiva per spingere la madre ad assistervi? Se è così, la sta ingannando, e sta facendo opera di manipolazione. Sta facendo in modo che la madre pensi qualcosa che non è vero (cioè che la ragazza sia devastata all'idea che sua madre non sia presente alla recita) e la manipola perché abbia una reazione emotiva che non è giustificata invece dalla situazione. Questo è il problema su cui attirano giustamente l'attenzione Barnhill e Noggle. Mostra che le persone possono essere manipolate in contesti che comportano decisioni propriamente non deliberative.

Se state decidendo se innamorarvi, potete essere manipolati – da una persona che vuole truffarvi o sedurvi. La vostra decisione, però, se è una decisione, può non essere, e forse non dovrebbe essere, una scelta riflessiva e deliberativa. O provate questo sentimento, oppure no. Se vi innamorate di una persona stupenda sotto tutti i punti di vista pertinenti, non è necessario che sia coinvolta una manipolazione, anche se il vostro innamoramento non comporta riflessione e deliberazione; provate a leggere *Possessione* di A.S. Byatt, per un esempio travolgente. Se però vi innamorate di una persona che sa come sollecitare tutti i vostri punti deboli, vi state facendo manipolare. Molta narrativa affronta questo problema: il romanzo di William Somerset Maugham del 1915, *Schiavo d'amore*, ne è un esempio formidabile.

Tutto questo è vero e importante. Se però vogliamo un bandolo per districare la matassa del problema della manipolazione, le definizioni (1)-(8) catturano gran parte del territorio pertinente, e la parte che catturano è di particolare rilievo per la politica e il diritto. Se teniamo presente la definizione (8), possiamo vedere una connessione essenziale fra l'idea della manipolazione e quella dell'inganno; possiamo addirittura vedere la prima come una versione più debole o più morbida della seconda. Chi inganna non rispetta la capacità di scelta riflessiva e deliberativa delle persone. Con

un atto di inganno, le persone quasi inevitabilmente[37] si sentono tradite e offese, non appena scoprono la verità. Lo stesso vale per la manipolazione: una volta che scopre il contesto completo, chi è stato manipolato in genere si sente tradito o usato. Si chiede: perché non mi è stato consentito decidere da me?

CASI ILLUSTRATIVI

Riflettiamo su qualche caso che sonda i confini del concetto di manipolazione.

1. *Le informazioni sul rischio relativo sono (fortemente) manipolative.* Supponiamo che i funzionari pubblici cerchino di persuadere le persone ad adottare determinati comportamenti, con l'aiuto di informazioni sul rischio relativo: "Se non fai x, le probabilità di morire per una patologia cardiaca triplicheranno!".[38] Supponiamo, però, che, per la popolazione pertinente, la probabilità di morte per patologie cardiache sia molto piccola (poniamo, 1 su 100.000) e che le persone siano molto più influenzate dall'idea di "triplicare il rischio" di quanto lo sarebbero se scoprissero che, se non fanno x, il rischio per loro passerebbe da 1/100.000 a 3/100.000. A dir poco, un aumento davvero modesto.

Oppure, supponiamo che l'affermazione pertinente arrivi da un medico, che cerca di convincervi ad assumere un certo farmaco. Il medico dice: "Questo farmaco può ridurre alla metà il tuo rischio annuo di infarto". Supponiamo che per voi il rischio di fondo sia 1 su 1000 e che, se assumete quel farmaco, il rischio si riduca a 1 su 2000. Supponiamo che il medico sappia benissimo che, esprimendo in quel modo la riduzione del rischio, aumentano le probabilità che assumiate il farmaco. Supponiamo che sappia benissimo anche che, se avesse detto che il farmaco può ridurre il vostro rischio da 1 su 1000 a 1 su 2000, si sarebbero potute verificare due cose. Innanzitutto, voi vi sareste sentiti confusi. In secondo luogo, avreste potuto pensare: "Uno su mille non sembra poi tanto male, e uno su duemila non mi sembra cambi un granché. Niente farmaco, grazie!".

Parlare di rischio relativo è un po' terrificante. Cattura l'attenzione di gran lunga più del rischio assoluto; il triplicarsi del rischio suona

allarmante, ma, se l'aumento è soltanto di 2 su 100.000, le persone potrebbero non preoccuparsi molto. Certamente è ragionevole fare la scelta di parlare di rischio relativo (che lascia immaginare un impatto forte sulla salute), nel tentativo di spaventare le persone e così manipolarle (almeno in un senso debole). Plausibilmente è una forma di raggiro che soddisfa la definizione (1), e certamente soddisfa la (2).

È vero che qualsiasi descrizione di un rischio impone delle scelte; chi descrive i rischi non può evitare qualche tipo di "effetto cornice" (*framing*), ma questo non è la stessa cosa della manipolazione. Si può correttamente argomentare che questa particolare scelta non rispetta le capacità deliberative delle persone; potrebbe addirittura essere un tentativo di puntare specificamente al Sistema 1. Come vedremo, questa conclusione non significa che l'uso della cornice del rischio relativo sia necessariamente fuori luogo.[39] Questo certo non è il caso più grave di manipolazione e, se dovesse salvare molte vite in una popolazione di grandi dimensioni, potrebbe essere giustificato. Ma non lo si può considerare manipolativo.

2. *L'uso dell'avversione alle perdite può essere manipolativo oppure no.* Supponiamo che i funzionari pubblici siano consapevoli del potere dell'avversione alle perdite,[40] cioè che le persone temono la prospettiva delle perdite molto più di quanto amino la prospettiva di guadagni equivalenti. "Perderai 1000 euro ogni anno se non applichi strategie di conservazione dell'energia" può essere più efficace di "Guadagnerai ogni anno 1000 euro se applichi strategie di conservazione dell'energia". Supponiamo che i funzionari utilizzino la "cornice della perdita", in modo da innescare nelle persone la preoccupazione per i rischi associati a un consumo eccessivo di energia o all'obesità. Potranno scegliere deliberatamente di sottolineare, in qualche tipo di campagna informativa, quanto le persone perderanno se non applicheranno tecniche di conservazione dell'energia, invece di quanto guadagneranno con l'applicazione di quelle tecniche.[41] L'uso dell'avversione alle perdite, con i suoi effetti potenzialmente significativi, è una forma di manipolazione?

La risposta non è ovvia, ma ci sono buone argomentazioni per sostenere che non sia una forma di manipolazione, perché le capacità deliberative delle persone sono trattate con rispetto. Anche nella cornice della perdita, le

persone restano pienamente in grado di valutare gli effetti complessivi. Non vengono ingannate. Bisogna però riconoscere che l'uso deliberato dell'avversione alle perdite può essere un tentativo di innescare i sentimenti negativi che sono tipicamente associati con le perdite.

È un problema diverso se l'uso dell'avversione alle perdite sollevi obiezioni etiche gravi. Nell'universo delle affermazioni verosimilmente manipolative, quelle che fanno ricorso all'avversione alle perdite non si può dire che rientrino fra le più problematiche, e, nel caso che stiamo esaminando, gli obiettivi del governo sono lodevoli. Se l'uso dell'avversione alle perdite produce grandi guadagni (in termini di salute o di benefici economici), non avremmo molti motivi di obiettare. Possiamo però identificare casi in cui l'uso dell'avversione alle perdite è interessato, e in cui il contesto lo rende un esempio genuino di manipolazione.[42]

Figura 1.1 Travel Assistance di IndiGo.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

Consideriamo, per esempio, i tentativi delle banche, a seguito di una nuova normativa della Federal Reserve Board, di sfruttare l'avversione alle perdite per spingere i clienti ad aderire a costosi programmi di protezione dagli scoperti, dicendo: "Non perdete la vostra protezione da scoperti di bancomat e carte di debito" e "RIMANETE PROTETTI con […] la copertura da scoperti di bancomat e carte di debito".[43] In casi simili, è evidente un tentativo di innescare un certo grado di allarme, quindi è ragionevole sostenere che i clienti vengono manipolati, a loro danno.

Come esempio meno vergognoso riferito a un'azienda che si rivolge ai consumatori, prendiamo il messaggio che IndiGo, una compagnia aerea indiana, presenta ai clienti accanto all'opzione di acquisto di un pacchetto aggiuntivo di assicurazione di viaggio (figura 1.1).

Questa non è la pratica commerciale peggiore che si possa immaginare, ma è giusto considerarla una forma di influenza che non rispetta la capacità di scelta riflessiva e deliberativa delle persone.

3. *L'uso di norme sociali non è necessariamente manipolativo.* Conoscendo la scienza comportamentale delle influenze sociali,[44] un pianificatore potrebbe considerare gli approcci seguenti.
   a) Informare le persone che la maggior parte degli individui nella loro comunità *mostrano un comportamento non desiderabile* (uso di droghe, abuso di alcol, evasione fiscale, azioni dannose per l'ambiente).
   b) Informare le persone che la maggior parte degli individui nella loro comunità *mostrano un comportamento desiderabile*.
   c) Informare le persone che la maggior parte degli individui nella loro comunità *è convinta che sia doveroso adottare un determinato comportamento*.

I primi due approcci si basano su "norme descrittive", cioè su quello che le persone fanno effettivamente.[45] Il terzo si basa su "norme ingiuntive", cioè su quello che le persone pensano che si debba fare. Sul piano empirico, si dà il caso che in genere le norme descrittive siano più potenti. Se quello che si cerca è un cambiamento nel comportamento, la cosa migliore è

sottolineare che la maggior parte delle persone effettivamente fa la cosa giusta. Se invece la maggior parte delle persone adotta il comportamento sbagliato, può essere utile invocare norme ingiuntive.[46]

Supponiamo che i funzionari pubblici siano ben consapevoli di questi risultati e che li utilizzino. Stanno compiendo atti di manipolazione? Senza fare molta violenza al linguaggio comune, qualcuno potrebbe pensare sia ragionevole concludere che è manipolativo scegliere la formulazione che avrà l'impatto maggiore. O, perlomeno, è così se le influenze sociali operano come fanno a causa del loro impatto sul sistema automatico, e se aggirano l'elaborazione deliberativa.[47] Sul piano empirico, la cosa è tutt'altro che chiara; le informazioni su ciò che le altre persone fanno o su ciò che pensano possono far parte della deliberazione riflessiva, e difficilmente vi si oppongono. Se i funzionari dicono il vero, accusarli di manipolazione significherebbe forzare i confini del concetto. Se le persone sono informate di ciò che fa la maggior parte della gente, i loro poteri di deliberazione sono rispettati. Non c'è raggiro.

4. *Le regole di default (predefinite) possono essere o non essere manipolative.* Le regole o le impostazioni predefinite (o, in campo giuridico, le norme dispositive) spesso rimangono attive, in parte per forza d'inerzia, in parte per il potere della suggestione.[48] Supponiamo che un funzionario pubblico sia consapevole di questo fatto e decida di riconsiderare una serie di regole predefinite per sfruttare questa loro "adesività". Per cercare di risparmiare denaro, potrebbe decidere per una impostazione predefinita per le stampanti, che preveda la stampa fronte-retro.[49] Cercando di ridurre l'inquinamento, potrebbe favorire, o addirittura richiedere, una norma dispositiva a favore delle energie rinnovabili.[50] Per far aumentare i risparmi, potrebbe promuovere, o addirittura imporre, l'adesione automatica a piani di pensionamento.[51]

Queste iniziative sono manipolative? Se le regole di default portano con sé un elemento di suggestione, una sorta di segnale informativo, non lo sono. Queste regole fanno appello a capacità deliberative e le rispettano, nella misura in cui veicolano informazioni su ciò che i pianificatori pensano che le persone dovrebbero acquistare. L'analisi è meno immediata se le regole predefinite rimangono attive per inerzia: senza una scelta cosciente,

le persone finiscono per essere iscritte a qualche tipo di programma o piano. In un certo senso, il funzionario pubblico sfrutta il Sistema 1, che è incline a inerzia e procrastinazione. La domanda è se l'iscrizione automatica manchi di rispettare la capacità di riflessione e deliberazione delle persone.

Per rispondere a questa domanda, è sicuramente pertinente il fatto che anche una regola di default che prevede il consenso esplicito (*opt-in*) è probabile rimanga altrettanto adesiva, e per gli stessi motivi, il che significa che la domanda è se *qualsiasi* regola di default vada considerata una forma di manipolazione. La risposta è semplice: non si può navigare nella vita senza regole di default e così, purché il funzionario pubblico non nasconda o sopprima qualcosa, e consenta di esercitare facilmente il consenso e il diritto di recesso, la scelta dell'una o dell'altra opzione non dovrebbe essere caratterizzata come manipolativa.[52]

Ci sono ancora motivi, comunque, per temere che le impostazioni predefinite a volte possano essere manipolative.[53] Consideriamo un altro caso. Un'azienda di social media adotta una norma di default, per la quale i suoi utenti devono rispettare vari termini relativi a privacy, newsfeed, reel e standard della comunità. La maggior parte di queste impostazioni può essere modificata, ma è necessario fare un certo sforzo. L'obiettivo della piattaforma è massimizzare il coinvolgimento degli utenti, che si traduce in profitti. Per quanto riguarda i reel, per esempio, per impostazione predefinita la piattaforma presenterà offerte che aumentino la probabilità di permanenza online degli utenti. Per qualcuno, i materiali offerti saranno legati a teorie della cospirazione; per altri, riguarderanno make-up e vestiti; per altri ancora, video sessualmente espliciti che si adattano ai loro gusti. Riflettendoci, molti utenti non sceglierebbero quelle offerte, anche se massimizzano il coinvolgimento. Riflettendo su ciò che è meglio, sceglierebbero qualcosa d'altro. Per impostazione predefinita, però, quelle sono le offerte che ricevono. Potremmo dire che le regole di default sono una sorta di inganno che non rispetta la capacità di scelta riflessiva e deliberativa delle persone.

Oppure, prendiamo un caso diverso. Se una persona noleggia online un'auto, per impostazione predefinita le verranno imposti una serie di condizioni e termini, fra cui assicurazioni di vario tipo, clausole che prevedono il pagamento di cifre elevate se l'auto non viene restituita con il pieno di carburante o la carica elettrica completa e una "assistenza autostradale" che si rivela molto costosa. Questi termini predefiniti non

hanno senso per ciò che devono fare molti utenti: costano troppo, ma, se vengono offerti per impostazione predefinita, i consumatori in genere li accettano. “Dal punto di vista di qualsiasi utente, la cosa migliore è trovare impostazioni predefinite che creino una situazione ottimale. Non è ragionevole però aspettarsi che le altre persone dispongano in quel modo le situazioni di scelta o addirittura che cerchino di farlo.”[54]

Termini predefiniti di questo tipo possono essere considerati manipolativi e lo sono certamente se non sono trasparenti e se chi deve scegliere non può modificarli facilmente. Ha senso considerare risposte regolative che forniscano informazioni e consentano di recedere facilmente, “con un solo clic”.

5. *L’organizzazione di un punto vendita può essere manipolativa.* Supponiamo che un supermercato organizzi i propri scaffali consapevolmente, con un unico obiettivo: massimizzare i ricavi. Supponiamo che collochi certi prodotti vicino alle casse, sapendo che le persone saranno particolarmente inclini a effettuare qualche acquisto d’impulso. Fra quei prodotti vi sono cibi e bevande molto calorici e ricchi di zuccheri. Supponiamo ancora che il supermercato collochi nei punti più in vista e ad altezza degli occhi i prodotti da cui estrae maggiori profitti, sperando di aumentarne la vendita. Supponiamo che il supermercato abbia accumulato una grande quantità di dati su quello che funziona e quello che invece non funziona. Spera di sfruttare attenzione e salienza in modo da aumentare le vendite e sa esattamente come farlo.

Il supermercato sta facendo opera di manipolazione? È giusto rispondere di sì. Usa varie tecniche in modo da ridurre la probabilità di una scelta riflessiva e deliberativa. È coinvolta la manipolazione. Se la valutazione dell’organizzazione (manipolativa) del punto vendita non è del tutto ovvia, le ragioni sono due. In primo luogo, nessun punto vendita può funzionare senza un criterio organizzativo; progettare la disposizione delle merci è inevitabile, e avrà un effetto (vedi il capitolo 7). In secondo luogo, certi modi di organizzare l’esposizione sono decisamente nell’interesse dei clienti. Possiamo discutere se un’organizzazione che massimizzi i ricavi cada in quella categoria. Comunque, può essere manipolativa.

6. *Affetto, attenzione, ancoraggio e contagio emotivo.* Gran parte della pubblicità moderna si rivolge al Sistema 1, con persone attraenti, colori vivaci e un'estetica caratteristica. (Pensate alle pubblicità del Viagra.) Spesso l'obiettivo è innescare un affetto particolare e più specificamente chiamare in causa la "euristica dell'affetto", per cui le persone valutano beni e attività facendo riferimento all'affetto che inducono – il che pone in forte rilievo il tema della manipolazione.[55]

L'euristica dell'affetto è stata identificata in origine grazie a un risultato degno di nota: quando viene chiesto loro di valutare i rischi e i benefici associati con determinate attività, le persone tendono a pensare che le attività rischiose offrano pochi benefici e le attività vantaggiose invece presentino rischi elevati.[56] In altre parole, le persone sono inclini a pensare che le attività che sembrano pericolose non portino benefici; è raro che considerino un'attività sia molto vantaggiosa sia molto pericolosa, oppure priva sia di vantaggi sia di pericoli. È una cosa estremamente strana. Perché le persone nella maggior parte dei casi non pensano che qualche attività possa essere sia molto vantaggiosa, sia molto rischiosa? Perché sembra che formulino una sorta di giudizio generale, di tipo Gestalt, che determina la valutazione di rischi e benefici? La riposta sembra essere che prima di tutto viene l'"affetto" e che questo contribuisca a orientare i giudizi sia sui rischi che sui benefici. Le persone, perciò, sono influenzate dall'euristica dell'affetto, per la quale hanno una reazione di tipo emotivo, "tutto considerato", di fronte a certi processi e prodotti. Quell'euristica funziona come una scorciatoia mentale per una valutazione più attenta. L'euristica dell'affetto lascia ampio spazio per la manipolazione. Si può indurre un affetto positivo o negativo, e influenzare tutti i tipi di giudizio, senza rispettare la capacità di scelta riflessiva e deliberativa delle persone.

Gran parte della progettazione di siti web è orientata ad attirare l'attenzione e a farla focalizzare nei punti giusti.[57] I gestori di telefonia mobile, i ristoranti e i negozi di abbigliamento utilizzano musica e colori in modo orientato a influenzare le reazioni affettive delle persone ai loro prodotti. Medici, amici e parenti (inclusi i/le partner) a volte fanno qualcosa di molto simile. I rapporti romantici sono un esercizio di manipolazione? Qualche volta la risposta è sicuramente sì, ma, se ci si chiede se le capacità deliberative delle persone vengano rispettate, in questo contesto sorge qualche difficoltà speciale.[58]

Gli avvocati difensori possono utilizzare la manipolazione; fa parte del loro lavoro, certamente quando si trovano di fronte a una giuria. Supponiamo che un avvocato sia a conoscenza del potere dell'"ancoraggio": quando alle persone si chiede di indicare un numero, qualsiasi numero venga portato a esempio, anche se arbitrario, può distorcere i loro giudizi. Chiedendo i danni per lesioni personali, un avvocato potrebbe citare il fatturato annuale di un'azienda (poniamo, un miliardo di euro), per produrre un'ancora. Questa può essere considerata una forma di manipolazione, perché non rispetta le capacità deliberative delle persone. Se un agente di vendita utilizza un'ancora per spingere le persone a pagare molto per un nuovo veicolo, anche qui è in campo la manipolazione, e per gli stessi motivi.

Lo stesso si può dire di alcune pratiche immaginabili nell'assistenza medica. Supponiamo che i medici vogliano che i pazienti scelgano particolari opzioni, e mettano in campo tecniche basate sulla scienza del comportamento per manipolarli affinché si orientino in quel senso. Un esempio potrebbe essere l'uso dell'euristica della disponibilità, quando le persone valutano i rischi in base a esempi che vengono facilmente in mente. Se un medico utilizza certi esempi ("Ho avuto un paziente che è morto quando…"), potremmo avere un caso di manipolazione, almeno se gli esempi sono utilizzati per aggirare o modificare giudizi deliberativi.

Oppure, torniamo a certi usi dei social media – come abbiamo visto, per esempio, nel 2012, quando Facebook ha cercato di orientare (manipolare) le emozioni di centinaia di migliaia di persone attraverso la presentazione di storie positive o negative, per vedere come influissero sul loro stato d'animo.[59] Molte pratiche delle piattaforme di social media, per quanto familiari, possono essere considerate manipolative nel senso pertinente.

UNA LISTA DI CONTROLLO

Anne Barnhill ha dedicato molto tempo e molto impegno a definire la manipolazione, ma è consapevole che i tentativi di definirla sono una bella sfida, e che forse non è possibile riuscire a dare una definizione su cui tutti siano d'accordo.[60] Invece di stare a preoccuparci se una pratica rientri o meno nella definizione, Barnhill pensa che dovremmo porci delle domande

che possano gettare luce sul perché qualcuno la considera manipolativa. Ecco alcune delle sue domande, leggermente modificate.

- L'influenza è occulta o nascosta?
- L'influenza prende a bersaglio o approfitta di debolezze o vulnerabilità psicologiche? Potrebbero essere debolezze o vulnerabilità della specifica persona presa di mira o delle persone in generale?
- L'influenza mina l'autocontrollo delle persone?
- L'influenza non coinvolge la deliberazione o la riflessione delle persone? Se è così, in questo contesto la deliberazione o la riflessione sono richieste?
- L'influenza fa sì che le persone abbiano reazioni emotive? Sono reazioni emotive appropriate o inappropriate, dato il contesto?
- L'influenza mina in qualche modo la razionalità o il ragionamento pratico?

Una lista di controllo di questo tipo è utile, perché è più facile affrontare queste domande, che non chiedersi se una certa affermazione o una certa azione vadano considerate manipolative. Con domande di questo tipo si possono affrontare molti casi di pattern (modelli di progettazione) oscuri, intesi come pratiche online ingannevoli o manipolative, anche se facciamo fatica a decidere se quei casi comportino manipolazione. Ci torneremo fra breve.

---

1. A.D.I. Kramer et al., "Experimental evidence of massive-scale emotional contagion through social networks", in *PNAS*, 111, 2014, p. 8788-8790, https://www.pnas.org/doi/10.1073/pnas.1320040111.
2. *Mad Men*, Quotes, IMDB, https://www.imdb.com/title/tt1105057/quotes.
3. "Nostalgia" significa in realtà "desiderio del ritorno a casa" e non "dolore di una vecchia ferita".
4. A. Barnhill, "How philosophy might contribute to the practical ethics of online manipulation", in F. Jongepier, M. Klink (a cura di), *The Philosophy of Online Manipulation*, Routledge, New York 2022, pp. 49, 50-51.
5. A. Isenberg, "Deontology and the ethics of lying", in *Philosophy and Phenomenological Research*, 24, 1964, pp. 463, 466.
6. T.L. Carson, *Lying and Deception*, Oxford University Press, Oxford 2010, p. 15.
7. J. Rawls, *A Brief Inquiry into the Meaning of Sin and Faith*, Harvard University Press, Cambridge (MA) 2009.

8. Vedi S. Cohen, “Manipulation and deception”, in *Australasia Journal of Philosophy*, 96, 2018, p. 483.

9. Vedi T.L. Carson, *op. cit.*, per una buona analisi del perché posizioni assolutiste o quasi assolutiste non funzionano; S. Bok, *Lying: Moral Choice in Public and Private Life*, Vintage Books, New York 1999, per una analisi di posizioni assolutiste o quasi assolutiste rispetto a menzogne e al mentire; vedi anche C.M. Korsgaard, “What’s wrong with lying?”, in J.E. Adler, C.Z. Elgin (a cura di), *Philosophical Inquiry: Classic and Contemporary Readings*, Hackett, Indianapolis 2007, per quella che si può considerare una condanna quasi assoluta delle menzogne, in base a motivazioni kantiane.

10. Vedi T. Nagel, “Concealment and exposure”, in *Philosophy and Public Affairs*, 27, 1998, p. 3.

11. Una rassegna eccellente è C. Coons, M. Weber (a cura di), *Manipulation: Theory and Practice*, Oxford University Press, Oxford 2014. Splendido è anche R. Noggle, *The Ethics of Manipulation*, 2018, in *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/ethics-manipulation.

12. R. Dworkin, *L’impero del diritto*, tr. it. il Saggiatore, Milano 1989.

13. T.M. Wilkinson, “Nudging and manipulation”, in *Political Studies*, 61, 2013, pp. 341, 342.

14. *Ibidem*, p. 145.

15. *Ibidem*. Una analisi particolarmente preziosa, che giunge a una conclusione diversa, si trova in A. Barnhill, “What is manipulation?”, in C. Coons, M. Weber (a cura di), *Manipulation: Theory and Practice*, cit., pp. 50, 72. Notate come, nel definire la manipolazione, si metta l’accento sul tentativo del manipolatore di influenzare scelte e decisioni. Se qualcuno cerca di mettere altre persone in un certo stato d’animo (per esempio, portandole a pranzo), non è coinvolta manipolazione se non viene compiuto alcun tentativo di influenzare le loro scelte, anche quando il tentativo di incidere sul loro stato d’animo non coinvolge un appello alle loro capacità di riflessione e deliberazione. (In effetti, i tentativi più efficaci di sollevare lo stato d’animo spesso non vi fanno appello.)

16. R. Faden, T. Beauchamp, *A History and Theory of Informed Consent*, Oxford University Press, Oxford 1986, pp. 354-368 (corsivo mio).

17. J. Raz, *The Morality of Freedom*, Clarendon Press, Oxford 1988, pp. 377-379.

18. A. Wood, “Coercion, manipulation, exploitation”, in C. Coons, M. Weber (a cura di), *Manipulation: Theory and Practice*, cit., pp. 17-18 (corsivo mio).

19. M. Gorin, “Toward a theory of interpersonal manipulation”, in C. Coons, M. Weber (a cura di), *Manipulation: Theory and Practice*, cit., pp. 73-97.

20. D. Susser, B. Roessler, H. Nissenbaum, “Technology, autonomy, and manipulation”, in *Internet Policy Review*, 8, giugno 2019, p. 4.

21. T.M. Wilkinson, *op. cit.*

22. Vedi P. Persson, “Attention manipulation and information overload”, in *Behavioral Public Policy*, 2, 2018, p. 78.

23. Ovviamente, alcune bugie sono giustificate; le intenzioni di chi mente possono essere meritorie (per esempio, per non urtare i sentimenti di qualcuno), e le conseguenze possono andare a sua discolpa (evitare danni gravi).

24. Per un’analisi rilevante nel contesto dell’inganno, vedi R. Craswell, “Interpreting deceptive advertising”, in *Boston University Law Review*, 65, 1985, p. 657; R. Craswell, “Regulating deceptive advertising: The role of cost-benefit analysis”, in *Southern California Law Review*, 64, 1991, p. 549.

25. R. Noggle, “Pressure, trickery, and a unified account of manipulation”, in *American Philosophical Quarterly*, 57, 2020, pp. 241, 242. Noggle offre un’analisi superba del raggiro, ma conclude che il termine non è abbastanza inclusivo. Concordo su questo aspetto, ma non penso invece che la pressione, su cui Noggle pone molto l’accento, sia una forma di manipolazione. Difendendo una spiegazione basata sull’errore, Noggle dice: “Sembra, dunque, che la pressione manipolativa possa avere successo solo quando il bersaglio sceglie erroneamente il bene minore. Di conseguenza, la pressione manipolativa e la manipolazione basata sul raggiro non sono, in fin dei

conti, tanto differenti. Sono entrambe tentativi di influenzare il comportamento del bersaglio facendo in modo che commetta un errore. La manipolazione basata sul raggiro comporta il tentativo di spingere il bersaglio ad adottare una credenza errata, un desiderio errato, un'emozione errata o qualche altro stato mentale errato. La pressione manipolativa cerca di spingere il bersaglio a rispondere acriticamente all'incentivo o disincentivo del manipolatore, come se il suo ordine di grandezza fosse molto maggiore di quel che è in realtà". Non discuterò in dettaglio questa interessante analisi, ma sarei cauto nel collegare l'idea di manipolazione con la produzione di errori, perché i manipolatori possono impedire gli errori, anziché indurli. Sarebbe possibile però intendere l'idea degli errori in modo che includa il processo, e non solo gli esiti, nel qual caso una "spiegazione basata sull'errore" può iniziare a convergere con la mia.

26. Vedi P. Persson, *op. cit.*

27. Vedi D. Kahneman, *Pensieri lenti e veloci*, tr. it. Mondadori, Milano 2012.

28. Vedi M. Nussbaum, *L'intelligenza delle emozioni*, tr. it. il Mulino, Bologna 2004.

29. C. Coons, M. Weber (a cura di), *op. cit.*, p. 16.

30. D. Kahneman, *op. cit.*; T. Wilson, *Strangers to Ourselves*, Harvard University Press, Cambridge (MA) 2014.

31. D. Kahneman, *op. cit.*

32. C. Camerer et al., "Neuroeconomics: How neuroscience can inform economics", in *Journal of Economic Literature*, 43, 1, 2005, p. 17.

33. R. Noggle, "Manipulative actions: A conceptual and moral analysis", in *American Philosophical Quarterly*, 34, 1995, p. 57.

34. A. Barnhill, *op. cit.*, p. 72. Barnhill sviluppa la superba trattazione pionieristica di Noggle, *op. cit*.

35. A. Barnhill, *op. cit.*

36. R. Noggle, *op. cit.* Noggle presenta le sue idee in dettaglio in *Manipulation: Its Nature, Mechanisms, and Moral Status*, Oxford University Press, Oxford 2025. Il suo è un libro straordinariamente valido, da cui ho imparato molto (come dal suo lavoro precedente). Sviluppa quella che definisce "spiegazione basata sull'errore" della manipolazione, che si riassume in questo modo: "La manipolazione è un'influenza che agisce inducendo il suo bersaglio a compiere un errore". Dovrebbe essere chiaro che non condivido questa spiegazione. La mia idea è che non esiste una spiegazione unificata della manipolazione e che quella su cui mi baso principalmente in questo capitolo, pensando a un tipo di raggiro che non rispetta la capacità delle persone di compiere scelte riflessive e deliberative, è essenziale ma incompleta. Noggle ha molto da dire a favore di quella che ha chiamato "spiegazione basata sul raggiro", che gli piace molto, ma crede che le sfugga quello che viene catturato da una "spiegazione basata sulla pressione". Ritengo inoltre che si possa dare manipolazione anche se porta le persone a decisioni corrette, anziché a errori; pensiamo a un medico che manipola un paziente perché si sottoponga a un'operazione che gli è necessaria. Nulla di quello che potrei dire qui comunque rende giustizia all'analisi di Noggle, a cui vi rimando, con ammirazione ed entusiasmo.

37. Possono essere grate, invece, se l'inganno è stato perpetrato consapevolmente per favorire i loro interessi.

38. T.M. Wilkinson, *op. cit.*, p. 347, usa questo esempio.

39. Vedi R. Craswell, "Regulating deceptive advertising…", cit., p. 552: "Le pubblicità sono 'prodotti' potenzialmente pericolosi e gli inserzionisti pubblicitari devono intraprendere passi ragionevoli per evitare che i consumatori vengano danneggiati dai loro prodotti. Gli inserzionisti però non dovrebbero essere biasimati se eventuali ulteriori cambiamenti avessero peggiorato, anziché migliorato, le cose per i consumatori".

40. Vedi E. Zamir, *Law, Psichology, and Morality: The Role of Loss Aversion*, Oxford University Press, Oxford 2015.

41. Vedi E. Aronson, *L'animale sociale*, tr. it. Maggioli, Milano 2019.

42. Vedi L.E. Willis, “When defaults pail: Slippery defaults”, in *University of Chicago Law Review*, 80, 2012, p. 1155.

43. *Ibidem*, p. 1192.

44. Per un riepilogo, vedi R.H. Thaler, C.R. Sunstein, *Nudge. La spinta gentile*, tr. it. Feltrinelli, Milano 2024.

45. Vedi R. Cialdini, “Crafting normative messages to protect the environment”, in *Currente Directions in Psychological Science*, 12, 2003, p. 105.

46. *Ibidem*; W. Schultz et al., “The constructive, destructive, and reconstructive power of social norms”, in *Psychological Science*, 18, 2007, p. 429.

47. Per risultati pertinenti (ma non decisivi), vedi C.J. Charpentier et al., “The brain’s temporal dynamics from a collective decision to individual action”, in *Journal of Neuroscience*, 34, 2014, p. 5816.

48. Vedi E. Johnson, D. Goldstein, “Decisions by default” in E. Shafir (a cura di), *The Behavioral Foundations of Public Policy*, Princeton University Press, Princeton 2013, p. 417; C.R. Sunstein, *Choosing Not To Choose*, Oxford University Press, Oxford 2015.

49. Vedi J. Egebark, M. Ekström, “Can indifference make the world greener?”, 2013, www2.me.su.se/paper/wp13_12.pdf.

50. Vedi C.R. Sunstein, L. Reisch, “Automatically green: Behavioral economics and environmental protection”, in *Harvard Environmental Law Review*, 38, 2014, p. 127.

51. S. Benartzi, R.H. Thaler, “Behavioral economics and the retirement savings crisis”, in *Science*, 339, 2013, p. 1152.

52. Vedi Z. Brown et al., “Testing the effects of defaults on the thermostat settings of OECD employees”, in *Energy Economics*, 39, 2013, p. 128; J. Beshears et al., *The Limitation of Defaults*, manoscritto non pubblicato, 15 settembre 2010, https://www.nber.org/sites/default/files/2023-06/orrc10-02-VD.pdf.

53. Vedi K. Grill, “Regulating online defaults”, in F. Jongepier, M. Klink (a cura di), *The Philosophy of Online Manipulation*, Routledge, New York 2022, p. 373.

54. *Ibidem*, p. 381.

55. Vedi P. Slovic (a cura di), *The Feeling of Risk: New Perspectives on Risk Perception*, Earthscan, London-New York 2010, pp. 3-20.

56. P. Slovic et al., “The affect heuristic”, in *European Journal of Operational Research*, 177, 2007, p. 1333; M. Finucane et al., “The affect heuristic in judgments of risks and benefits”, in *Journal of Behavioral Decision Making*, 13, 2000, p. 1.

57. S. Krug, *Don’t make me think. Un approccio di buon senso all’usabilità web e mobile*, tr. it. Tecniche Nuove, Milano 2014.

58. Non conosco un’analisi dettagliata di questo aspetto, ma, per una discussione pertinente, vedi E. Cave, “Unsavory seduction and manipulation”, in C. Coons, M. Weber (a cura di), *Manipulation: Theory and Practice*, cit., p. 176.

59. Vedi A.D.I. Kramer et al., *op. cit.*

60. Vedi A. Barnhill, *op. cit.*

2

VALUTARE LA MANIPOLAZIONE

Quando la manipolazione è sbagliata? Perché è sbagliata?

Alcune forme di manipolazione fanno solo sorridere e scuotere la testa. Se la vostra partner vi chiede di fare qualche lavoretto in casa e aggiunge, con aria seducente, “Ti prometto che ne sarà valsa la pena”, potreste sentirvi manipolati, ma potrebbe non importarvi molto. Se una pubblicità politica utilizza musica allegra e brillante per un candidato e musica sgradevole e stridente per il suo avversario, potreste pensare: “Funziona sempre così, con la politica”. Se una pubblicità per una nuova automobile mostra persone attraenti che vanno a una festa elegante, potreste dire: “Ho già capito che cosa vogliono farmi pensare”.

A volte però la manipolazione produce una reazione molto più negativa. Ci si sente offesi o umiliati. Si può percepire la manipolazione come una forma di furto. Qualcuno ci sta rubando qualcosa. In effetti, questa è una componente importante del quadro.

R-I-S-P-E-T-T-O

Il problema più ovvio della manipolazione è che può insultare sia l’autonomia che la dignità. Un atto di manipolazione non tratta con rispetto le persone. Dal punto di vista dell’autonomia, il problema è che la manipolazione può privare i soggetti persone della capacità di agire autonomamente; è una parente stretta della coercizione. Se una persona viene manipolata in modo da acquistare un prodotto, può addirittura sentirsi costretta con la forza.

Dal punto di vista della dignità, un problema è che la manipolazione può essere umiliante. Adulti sani, che non soffrono di una mancanza di capacità, non devono essere tratti in inganno; devono essere trattati come persone pienamente in grado di prendere le proprie decisioni. La loro autorità sulla propria vita non deve essere minata da approcci che le trattino come bambini o come marionette. Come dice Marcia Baron, "la persona virtuosa cerca di ragionare con l'altro, non lo lusinga né lo inganna perché agisca in modo diverso […] essere manipolativi è un vizio, per la sua arroganza e presunzione, e perché la persona manipolativa ricorre fin troppo facilmente a trucchi […]".[1]

Supponiamo, per esempio, che qualcuno pensi: "Voglio che tutti i miei amici facciano certe cose, e conosco varie strategie per fare in modo che le facciano. Ho letto molto a proposito di psicologia e scienza del comportamento, comprese le opere migliori sull'influenza sociale, e il mio progetto è utilizzare quello che so per manipolare i miei amici".[2] Una persona del genere non rispetterebbe l'autonomia dei suoi amici. Li userebbe come degli strumenti. Le sue azioni, in effetti, non sarebbero coerenti con la natura stessa dell'amicizia, che comporta una relazione che non sia strettamente o prevalentemente strumentale.

Veniamo al caso del governo. Supponiamo che funzionari pubblici (poniamo, nell'ufficio di un presidente) imparino analogamente molto su come influenzare le persone, e supponiamo che decidano di utilizzare quello che hanno appreso per raggiungere certi obiettivi politici. Supponiamo che alcuni degli strumenti in questione cerchino di stravolgere o aggirare la deliberazione. Dobbiamo sapere i dettagli – Che cosa, esattamente, fanno? –, ma probabilmente è giusto dire che è coinvolta la manipolazione, e che quei funzionari non rispettano a sufficienza l'autonomia dei loro cittadini. Dobbiamo ancora conoscere i dettagli, ma saremmo giustificati anche a dire che quei funzionari non trattano con rispetto i cittadini; è possibile che li stiano usando come strumenti o marionette per i loro scopi (o magari per le finalità di servizio pubblico che preferiscono).

Sotto questa luce, dovremmo riuscire a vedere che, nella valutazione della manipolazione, il *ruolo* ha una grande importanza. Supponiamo che Jones stia cercando un lavoro. Non è certo inaccettabile che Jones cerchi di fare una buona impressione sui potenziali datori di lavoro e, se studia la scienza del comportamento e come influenzare le persone, è difficile

pensare che sarebbe illegittimo che sfruttasse quello che ha appreso. Certo, esistono limiti etici a quello che Jones può fare e, anche per qualcuno che sta cercando un lavoro, le forme più smaccate di manipolazione passerebbero il segno. In interazioni o relazioni che sono strumentali, e che sono intese in questo senso, però, alcuni dei vincoli etici sono allentati o sono almeno differenti.

In una campagna pubblicitaria, tutti sanno qual è la natura dell'interazione. La manipolazione è la moneta corrente del regno. Lo scopo delle pubblicità è vendere prodotti e, anche se si possono trovare presentazioni puramente fattuali, molte pubblicità non si rivolgono affatto alla riflessione o alla deliberazione: cercano invece di creare certi stati d'animo e certe associazioni. Qualcosa di simile si può dire per alcuni aspetti delle campagne politiche. La relazione fra una campagna e gli elettori ha un carattere strumentale: le campagne vogliono voti, e la cosa è chiara a tutti. Nel processo, sia le pubblicità sia i comizi avranno caratteristiche manipolative. Certo, sarebbe un po' stravagante dire che, in quei casi, le persone hanno dato il loro consenso alla manipolazione in tutte le forme immaginabili. Anche qui è possibile che vengano valicati dei confini. (Le caselle preselezionate possono valicarli.) È importante però che le persone siano consapevoli della natura peculiare delle iniziative pertinenti.

Altri ruoli sono accompagnati da norme diverse e la manipolazione può non essere coerente con quelle norme, o addirittura violarle. Quando i governi si rivolgono ai loro cittadini, le norme sono radicalmente diverse da quelle che si applicano alle campagne elettorali, o almeno è così nelle società libere e democratiche, in cui si intende che la sovranità ultima spetta al pubblico. Certo, non si può dire che ai funzionari sia proibito presentare le opzioni in modo tale che quelle che preferiscono siano poste sotto la luce più favorevole. Se le caratteristiche manipolative delle loro azioni diventano più estreme, diventa più ampio anche lo spazio di legittimità delle obiezioni.

BENESSERE

Supponiamo di essere "welfaristi", cioè di credere che ciò che conta è come va la vita delle persone.[3] Vogliamo sapere se le persone sono tristi o

felici, calme o tese, rilassate o spaventate, stressate o piene di gioia. Supponiamo inoltre che ci preoccupino le violazioni di autonomia e dignità, nella misura in cui influiscono sulle esperienze soggettive delle persone (facendole sentire, per esempio, in collera o umiliate). Se le cose stanno così, che cosa dovremmo pensare della manipolazione?

Dovrebbe essere chiaro che *non esiste tabù per la pratica*. I welfaristi sono disposti ad accettare menzogne e inganno, se miglioreranno la vita delle persone, e lo stesso vale per la manipolazione, che può promuovere il benessere delle persone. Se un medico manipola una paziente, così che assuma un farmaco che le salverà la vita, potremmo non avere alcun motivo di obiettare. Ciononostante, esiste una caratteristica obiezione welfarista alla manipolazione, che assume la forma seguente. Almeno gran parte delle volte, chi deve fare una scelta sa che cosa è nel suo migliore interesse (almeno se si tratta di persone adulte che non hanno un problema di capacità).[4] Sa di preferire il tennis al bowling, o i cani ai gatti, o Taylor Swift a Drake, o le commedie romantiche ai film dell'orrore. Ha un accesso esclusivo alle proprie situazioni, ai propri limiti e ai propri gusti. Se viene manipolato, è privato della (completa) capacità di fare scelte proprie, semplicemente perché non gli viene concessa una possibilità equa o adeguata di pesare tutte le variabili. Se qualcuno vuole aiutare le persone a compiere scelte migliori, ha l'obbligo di informarle, in modo che possano fare le proprie valutazioni da sé.

Il problema del manipolatore è che non possiede certe conoscenze pertinenti – sulla situazione, i gusti e i valori di chi deve effettuare la scelta. In assenza di quelle conoscenze, sovverte comunque il processo grazie al quale chi deve scegliere prende le proprie decisioni in merito a ciò che è meglio per sé. Le cose peggiorano ulteriormente se il manipolatore ha in mente solo i propri interessi, anziché quelli di chi deve scegliere. In questo senso si può dire che un manipolatore interessato solo a se stesso ruba alle persone, sia limitando la loro capacità di agire, sia spostando le loro risorse nella direzione che preferisce.

Per questi motivi, l'obiezione del welfarista al paternalismo ha le sue radici nelle stesse preoccupazioni che stanno alla base del "principio del danno" di John Stuart Mill. Nella concezione di Mill, il problema degli outsider, compresi i funzionari governativi, è che non possiedono le informazioni necessarie. Mill insiste che l'individuo è "la persona più interessata al proprio benessere",[5] e che "l'uomo e la donna più comuni

hanno strumenti di conoscenza incommensurabilmente superiori a quelli di chiunque altro". Quando la società cerca di prevalere sul giudizio dell'individuo, lo fa sulla base di "supposizioni di carattere generale", che "possono essere del tutto errate e, ammesso siano giuste, è altrettanto possibile che vengano applicate male ai casi individuali". Queste osservazioni valgono per chi applica la manipolazione, non meno che a chi esercita la coercizione.

Nonostante ciò, dovrebbe essere chiaro che, dal punto di vista del welfarista, non dovrebbe esserci un divieto della manipolazione.[6] Tutto dipende dal fatto che la manipolazione migliori il benessere delle persone. Un manipolatore può non rispettare la capacità di scelta riflessiva e deliberativa delle persone, ma ciononostante migliorare il loro benessere.

Per capirci, immaginate un manipolatore benigno, onnisciente, che promuove il benessere (una sorta di genitore idealizzato) e si preoccupa solo del benessere di coloro che vengono manipolati, che possiede tutte le conoscenze di cui ha bisogno, e che semplicemente non commette mai errori. In base all'ipotesi, il manipolatore che promuove il benessere dovrebbe essere apprezzato, secondo i criteri welfaristi.[7] L'aspetto principale è che, se le persone sanno di essere manipolate, e a loro non sta bene, ci sarà una perdita di benessere e quella perdita dovrà essere tenuta in conto nella valutazione complessiva. Il punto semplicemente è che, se le persone odiano i manipolatori, è meno probabile che la manipolazione sia sopportabile in base a criteri di benessere (a meno che sia occulta, il che fa sorgere problemi suoi).[8]

Mettiamo da parte questo punto. Il problema principale dell'esperimento ideale è che è improbabile che i manipolatori siano benigni o onniscienti. Di solito hanno i loro programmi e il fatto che si dedichino alla manipolazione lo attesta. Se avessero genuinamente a cuore il benessere di chi deve scegliere, perché non cercare di persuaderlo? Perché superare il confine della manipolazione? Certo, il manipolatore può essere in grado di rispondere a questa domanda se, per esempio, la tempestività è essenziale, o se chi deve scegliere non ne ha la capacità (perché, per esempio, è un bambino o è gravemente malato). Oppure supponiamo che avvertimenti basati su immagini choc sui pericoli per la salute, che manipolano le persone, salvino molte vite; supponiamo anche che non si possano salvare molte vite con una presentazione puramente fattuale, non accompagnata da

quelle immagini. In base a criteri welfaristi, si potrebbe dire molto a favore dell'uso di immagini choc come avvertimento sui pericoli per la salute.[9]

L'esempio mostra che, dal punto di vista del benessere, tutto dipende dal contesto; il fatto che sia coinvolta la manipolazione non incrimina *necessariamente* il calcolo del benessere da parte del manipolatore. In molte situazioni, però, il sospetto sugli obiettivi dei manipolatori è perfettamente giustificato. A questo punto bisogna aggiungere che, anche quando quegli obiettivi sono ammirevoli, è possibile che i manipolatori non abbiano conoscenze sufficienti per giustificare le loro azioni. Consideriamo la notevole osservazione di Friedrich Hayek, secondo cui "la consapevolezza della nostra irrimediabile ignoranza della maggior parte di ciò che è noto a qualcuno [che deve compiere una scelta] è *la base principale dell'argomento a favore della libertà*".[10] Tornerò, al momento opportuno, su questo punto e sulla questione della regolamentazione.

CONSENSO

Supponiamo che le persone diano il loro consenso alla manipolazione.[11] Un alcolista potrebbe dire a sua moglie: "Voglio assolutamente smettere. Per favore, usa qualsiasi tecnica ti venga in mente per aiutarmi. Non ho nulla contro la manipolazione, anzi è la benvenuta!". Oppure supponiamo che la stragrande maggioranza dei fumatori dica al governo: "Voglio proprio smettere! Se potete trovare un modo per aiutarmi a superare la mia dipendenza, ve ne sarei grato". T.M. Wilkinson nota che è troppo rozzo dire che la manipolazione viola l'autonomia, perché "si può dare il proprio consenso alla manipolazione. Se è stato dato il consenso, nel modo giusto, allora la manipolazione sarebbe almeno coerente con l'autonomia e si potrebbe addirittura dire che la rafforzi".[12]

La conclusione ha molta forza. La vicenda di Ulisse e le sirene è istruttiva qui, indipendentemente dal fatto che Ulisse chiedesse coercizione, manipolazione o qualche altra cosa.[13] In effetti, chiedeva coercizione:

> Delle Sirene dal canto divino per prima cosa ordinava
> che fuggissimo e voce e prato fiorito.
> A me solo ordinava d'udire quel canto; ma voi con legami
> strettissimi dovete legarmi, perché io resti fermo,
> in piedi sulla scarpa dell'albero: a questo le corde m'attacchino.

E se vi pregassi, se v’ordinassi di sciogliermi,
voi con nodi più numerosi stringetemi.

(*Odissea*, XII, 158-164, tr. it. Calzecchi Onesti)

Non c’è neanche un’obiezione ovvia, nel caso del consenso, dal punto di vista del benessere. Chi sceglie ha deciso che le cose gli andranno meglio se viene manipolato. Se pensa che la sua scelta presumibilmente promuoverà il suo benessere, dobbiamo rispettarla, anche se ciò che ha scelto è la manipolazione.

Nei casi più facili, il consenso è esplicito. In quelli più difficili è solo implicito, nel senso che il manipolatore lo deduce dalle circostanze o crede, con validi motivi, che chi deve fare la scelta darebbe il proprio consenso, se gli venisse chiesto. Se l’inferenza è meno che affidabile, la giustificazione del consenso è analogamente indebolita. Se la convinzione è ragionevole ma potenzialmente sbagliata, avrebbe senso insistere sull’ottenere un consenso esplicito, per evitare il rischio di errore. È importante vedere che la manipolazione consensuale è un caso inusuale; chi ha bisogno di aiuto di solito non chiede: “Per favore, manipolami”. Casi del genere però certamente esistono, almeno quando le persone hanno problemi seri di autocontrollo.

TRASPARENZA

A volte si assume che l’idea della manipolazione implichi una mancanza di trasparenza, come se qualcosa di importante venisse nascosto o non divulgato.[14] Se un manipolatore si comporta come un burattinaio, potrebbe tacere il proprio ruolo; questa sarebbe una caratteristica importante della manipolazione. Rispetto alla manipolazione, però, non è del tutto chiaro che cosa significhi “trasparenza”.

A volte la manipolazione stessa *consiste* in una mancanza di trasparenza in merito a un aspetto pertinente di una situazione; è in campo allora la vera manipolazione. (Un genitore dice a un bambino piccolo: “Se sarai davvero bravo, Babbo Natale ti porterà una giraffa di peluche”.) Quando l’aspetto pertinente è rivelato, la manipolazione sparisce. Il carattere manipolativo di alcune azioni però non consiste nell’essere occulta. (Ricordate l’uso della riduzione relativa del rischio.) Alcune azioni possono essere sia manipolative, sia pienamente rivelate a chi viene manipolato.[15]

Un’immagine che costituisca un avvertimento sui pericoli per la salute, per esempio, è perfettamente trasparente (e, se è obbligatoria per una norma, può essere addirittura preceduta da un periodo per i commenti del pubblico). La pubblicità subliminale potrebbe essere preceduta da un avviso esplicito: “Questo film contiene pubblicità subliminali”.

Nella scena cruciale de *Il mago di Oz*, il mago dice di non prestare attenzione all’uomo dietro la tenda: l’uomo dietro la tenda, ovviamente, è semplicemente un essere umano travestito da Grande Mago, che pretende di avere un’autorità di gran lunga superiore a quella che merita e definisce le situazioni sociali in un modo che nasconde caratteristiche che, se rivelate, modificherebbero molto i giudizi e le scelte delle persone. Oppure pensiamo a un film magnifico ma meno celebrato, *The Truman Show*, in cui ciò che avviene nella vita di Truman, l’eroe-protagonista del film, è influenzato dalle molte decisioni prese da un grande manipolatore, che nasconde il fatto che Truman sia la stella involontaria di uno spettacolo televisivo e che i suoi amici e conoscenti non sono altro che attori. Nascondere e occultare sono forme di inganno, e sono caratteristiche comuni della manipolazione. Ogni volta che le persone che impongono la propria influenza nascondono il proprio ruolo, sembra ragionevole obiettare. Una mancanza di trasparenza offende sia l’autonomia, sia la dignità.

Dal punto di vista del benessere, potremmo chiederci perché, esattamente, qualcuno non sia stato onesto con chi deve compiere la scelta, che dovrebbe essere in grado di prendere le proprie decisioni, armato delle informazioni pertinenti. Come prima, però, l’analisi del benessere è più complessa, perché possiamo immaginare casi in cui la trasparenza non è necessaria e, in effetti, può essere un problema. Supponiamo che qualcuno soffra di un grave disturbo dell’autocontrollo e che la sua vita sia in pericolo (a causa, poniamo, della dipendenza da alcol o stupefacenti). Supponiamo anche che un manipolatore abbia individuato una strategia in grado di salvargli la vita e che la trasparenza renderebbe meno efficace la manipolazione. Per ipotesi, le considerazioni welfariste sono contrarie alla trasparenza. La forza intuitiva di argomenti di questo tipo è massima quando si applicano a persone che non hanno capacità (bambini piccoli, chi soffre di disturbi mentali), ma possiamo immaginare anche contesti in cui persone adulte in pieno possesso delle loro capacità trarrebbero beneficio dall’essere manipolate.[16]

Può esistere una giustificazione welfarista anche per la manipolazione nascosta in altre circostanze estreme – per esempio, quando si cerca di impedire un sequestro di persona o di salvare una vittima di sequestro. Se l'obiettivo è fermare un malfattore, o qualcuno che minaccia di fare realmente del male, può essere perfettamente accettabile o addirittura obbligatorio manipolare quelle persone, e tenere nascosto questo fatto. Quelle persone hanno rinunciato al loro diritto di essere trattate con rispetto e il loro benessere, in quanto individui capaci di scelta, non è una questione che ci stia molto a cuore.

In casi standard, però, questo argomento non sarà disponibile. Ne segue che, nella maggior parte delle occasioni, la manipolazione non deve essere nascosta o occulta, anche quando è giustificata; torniamo al caso delle immagini esplicite come avviso di pericolo (che possono essere considerate manipolative o meno). La trasparenza è una condizione necessaria. Notate, però, che non è sufficiente. La pubblicità subliminale non diventerebbe più accettabile semplicemente perché le persone ne sarebbero state informate. Se una catena di sale cinematografiche annunciasse che le sue anteprime saranno piene di pubblicità subliminali, le persone potrebbero giustamente disapprovare.

Come esempio deplorevole di mancanza di trasparenza, pensate all'offerta di un abbonamento mensile al sito online di Eric July, musicista e figura di spicco nel campo dei media (figura 2.1). Ai clienti viene offerto un abbonamento per 99,99 dollari al mese, ma l'informazione che l'abbonamento sia "mensile" è scritta nello stesso colore dello sfondo della pagina web. Il fatto che i clienti stiano per sottoscrivere un canone di abbonamento che verrà addebitato ogni mese non è affatto trasparente, e lo scarso livello di trasparenza fa pensare che si tratti di un caso di manipolazione.

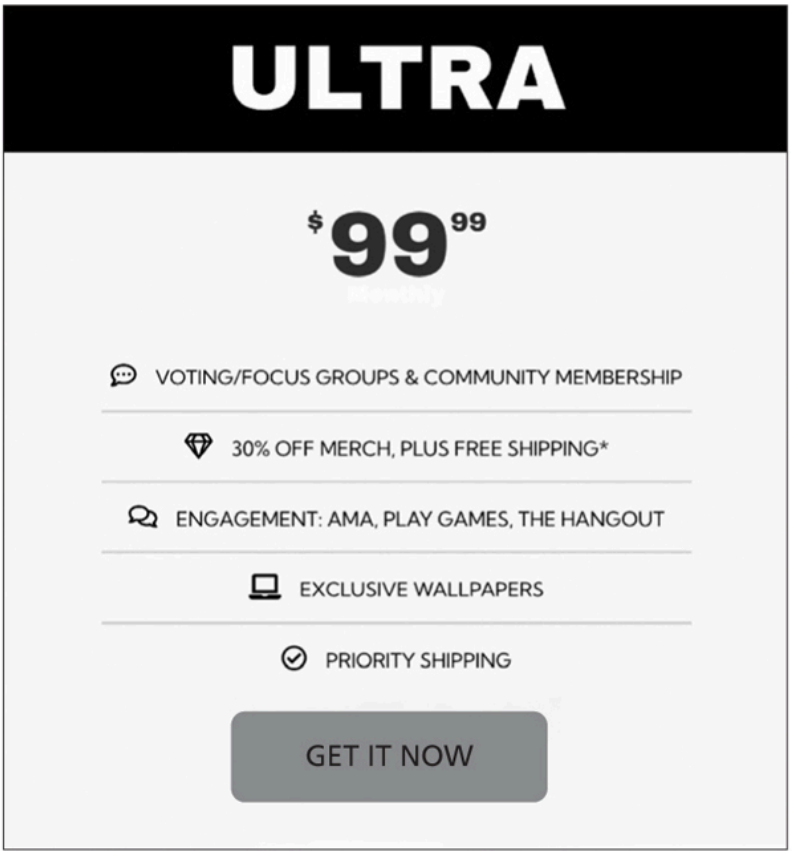

Figura 2.1 Abbonamento online al sito di Eric July.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

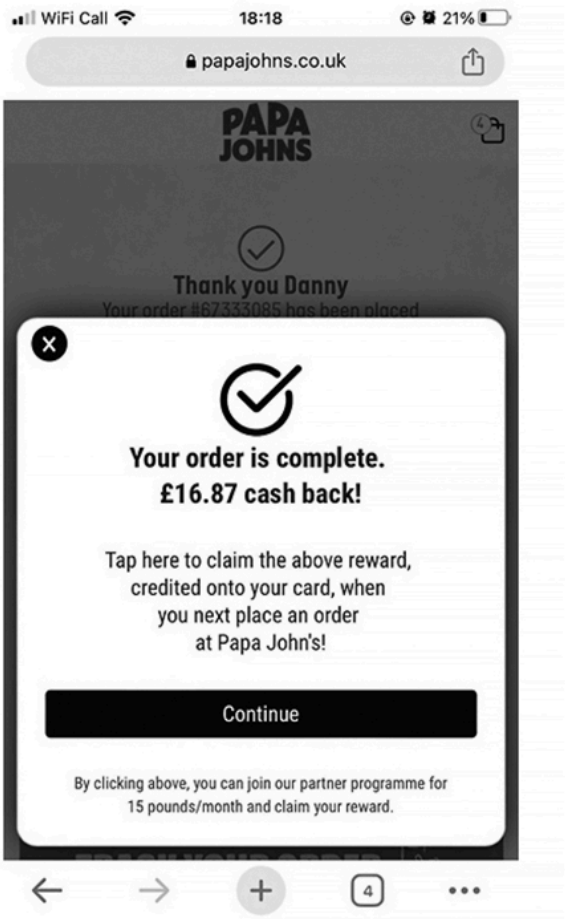

Figura 2.2 Il cashback di Papa John.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

I problemi di trasparenza si possono presentare con gradi diversi di gravità. Consideriamo l'offerta di cashback di Papa John (figura 2.2). L'offerta dice che, accettando le 16,87 sterline di cashback, il cliente acconsente a un abbonamento che costa 15 sterline al mese. La condizione, comunque, non è del tutto trasparente. Che cosa è più probabile che l'utente noti? L'abbonamento e il suo costo sono in caratteri di corpo deplorevolmente piccolo.

AUTORIZZAZIONE DEMOCRATICA

E se la manipolazione è autorizzata democraticamente? Supponiamo che un Parlamento nazionale voti espressamente a favore, magari al fine di

migliorare la salute pubblica (per esempio, per scoraggiare il fumo o abitudini alimentari poco sane) o magari per promuovere altre finalità (per esempio, aumentare la partecipazione alle consultazioni elettorali o il patriottismo). In casi relativamente benigni, in cui la manipolazione coinvolta è minima o nulla, i legislatori possono sostenere una campagna educativa pensata per ridurre le malattie e i decessi e che comporti una serie di strategie, basate sulle scienze del comportamento, per realizzare gli obiettivi.

Dovrebbe essere chiaro che l'autorizzazione democratica di per sé non dovrebbe eliminare obiezioni altrimenti convincenti alla manipolazione. I problemi più ovvi sorgono se i legislatori hanno finalità non legittime (poniamo, inculcare pregiudizi razziali o consolidare in qualche modo la propria posizione). Le obiezioni ben note (che riguardano autonomia, dignità e benessere) valgono anche se i fini sono legittimi. Se un Parlamento nazionale autorizza la pubblicità subliminale, rimane del tutto possibile opporsi sulla base sia dell'autonomia, sia della dignità. È possibile un'obiezione anche dal punto di vista del benessere: perché il processo democratico ha autorizzato la manipolazione, anziché qualche altro tipo di comunicazione? Qualche altro tipo di comunicazione avrebbe avuto maggiori probabilità di aumentare il benessere?

Certo, potremmo intendere l'autorizzazione democratica come una forma di consenso collettivo o a maggioranza; le persone potrebbero vedere di buon occhio un po' di manipolazione (o magari anche molta). Perché però a una maggioranza dovrebbe essere permesso di consentire la manipolazione di tutti? Esistono anche rischi evidenti nell'autorizzare i funzionari pubblici a perseguire questa linea di argomentazione. L'obiezione alla manipolazione viene da individui che non vogliono essere manipolati; il fatto che una maggioranza voglia manipolarli non la rende più accettabile.

In certi contesti, però, l'argomento a favore di almeno un modesto grado di manipolazione può non essere del tutto implausibile, e appare ancora più forte se il processo democratico lo ha sostenuto. Immaginiamo, per esempio, una campagna di educazione pubblica pensata per ridurre i rischi associati con l'uso del cellulare mentre si guida, o un'iniziativa volta a combattere l'uso di sostanze pericolose o a convincere le persone a non abbandonare la scuola. Molte campagne di questo genere sono intense e hanno una componente emotiva; possono essere viste come iniziative volte

a combattere problemi di autocontrollo e a fare sì che le persone si concentrino sul lungo periodo. Possono rispettare o non rispettare la capacità di scelta riflessiva e deliberativa dei singoli. Dovrebbero rispettarla, ma, sulla base di ipotesi immaginabili in merito ai loro effetti benefici, forse non dovremmo preoccuparcene molto.

Se il governo manipola la gente, magari attraverso appelli emotivamente evocativi, potrebbe anche rispondere al fatto che le persone sono state già manipolate, e a loro danno. Se si parla di sigarette, per esempio, è plausibile dire che manipolazioni precedenti (fra cui pubblicità e norme sociali) hanno influenzato le persone e le hanno fatte diventare fumatrici. Se è così, forse possiamo dire che ai funzionari pubblici è permesso combattere il fuoco con il fuoco. È ragionevole però sostenere che due torti non fanno una ragione – e che, se il governo cerca di spingere le persone a smettere di fumare, deve trattarle come adulte e fare appello alle loro capacità deliberative.

Bisogna ricordare che la manipolazione si presenta in gradi diversi, ed esiste una grande differenza fra una bugia e un tentativo di presentare un'alternativa sotto una luce attraente, scostante o orribile. Nella vita comune, non è probabile che accusiamo di manipolazione gli amici o le persone a cui vogliamo bene, se hanno caratterizzato un certo comportamento come quello preferito dalla maggior parte dei membri del gruppo dei nostri pari, o se hanno sottolineato le perdite che potrebbero accompagnare un'alternativa che disapprovano fortemente, o se avessero accompagnato la descrizione di un'opzione con aria grave e corrucciata. Abbiamo visto che esistono forme al massimo deboli di manipolazione (se di manipolazione si può parlare), ed è importante vedere che quelle forme deboli possono essere anche accettabili e benigne (e magari un po' divertenti) se promuovono gli interessi delle persone verso le quali sono rivolte.

---

1. Vedi M. Baron, "Manipulativeness", in *Proceedings and Addresses of the American Philosophical Association*, 77, 2003, p. 37. Più in generale, questo il suo suggerimento: "Al contrario, la persona che possiede la virtù corrispondente alla manipolatività – una virtù per la quale, credo, non abbiamo un nome – sa quando sia appropriato cercare di provocare un cambiamento nella condotta di altri e lo fa per le ragioni giuste, per gli scopi giusti, e solo quando è giustificato (e vale la pena correre i rischi che comporta) e solo utilizzando mezzi accettabili. La persona virtuosa cerca di ragionare con l'altro, non lo lusinga né lo inganna perché agisca in modo diverso […] essere

manipolativi è un vizio, per la sua arroganza e presunzione, e perché la persona manipolativa ricorre fin troppo facilmente a trucchi […]" (*ibidem*, pp. 48, 50).

2. Una fonte classica in proposito è R. Cialdini, *Le armi della persuasione*, tr. it. Giunti, Firenze 2017. Un altro classico, molto meno accademico ma altamente informativo, è D. Carnegie, *Come trattare gli altri e farseli amici*, tr. it. Bompiani, Milano 2020. Entrambi i libri presentano molte indicazioni sulla manipolazione efficace.

3. Metto tra parentesi qui varie domande sul modo migliore di intendere il welfarismo. È possibile avere una concezione del benessere che tenga conto anche di autonomia e dignità. Vedi A. Sen, *Lo sviluppo è libertà*, tr. it. Mondadori, Milano 2020; M. Nussbaum, *Creare capacità*, tr. it. il Mulino, Bologna 2014; A. Sen, "Utilitarianism and welfarism", in *Journal of Philosophy*, 76, 1979, p. 463. Per un'analisi molto istruttiva, vedi M. Adler, *Well-Being and Fair Distribution*, Oxford University Press, Oxford 2011.

4. Un bambino, o una persona che soffra di qualche forma di demenza, può opporsi solo debolmente alla manipolazione. I genitori manipolano regolarmente i figli piccoli, in parte per favorire il loro benessere. Le persone che si prendono cura di loro manipolano le persone che soffrono di demenza. Queste pratiche sono date in gran parte per scontate, ma potremmo immaginare situazioni in cui sollevino gravi questioni etiche. Anche se la manipolazione pertinente è nell'interesse di chi viene manipolato, l'interesse per l'autonomia e la dignità impone dei vincoli anche qui.

5. J. S. Mill, *Sulla libertà*, tr. it. Bompiani, Milano 2000, p. 239.

6. Va notato che la domanda qui è se la manipolazione aumenti o diminuisca il benessere, non se la legge o qualche regolatore debba metterla al bando. La seconda domanda solleva problemi di competenza istituzionale e di costi di decisione; inoltre richiede attenzione all'effetto della manipolazione su grandi popolazioni con modi d'intendere eterogenei. In risposta a una campagna pubblicitaria, per esempio, alcune persone possono essere manipolate (nel senso che essenzialmente è solo il loro Sistema 1 che viene influenzato), mentre altre no (perché la campagna innesca una quantità significativa di deliberazione).

7. Cfr. la difesa vigorosa del paternalismo coercitivo in S. Conly, *Against Autonomy*, Cambridge University Press, Cambridge (UK) 2012.

8. La manipolazione occulta è rischiosa, perché potrebbe essere smascherata, e le persone non saranno felici nello scoprire di essere state manipolate occultamente.

9. C. Jolls, "Product warnings, debiasing, and free speech: The case of tobacco regulation", in *Journal of Institutional and Theoretical Economics*, 169, 2013, p. 53.

10. Vedi F.A. von Hayek, "The market and other orders", in B. Caldwell (a cura di), *The Collected Works of F.A. Hayek*, Chicago University Press, Chicago 2013, p. 384 (corsivo mio).

11. Secondo alcuni, il concetto di manipolazione presuppone un'assenza di consenso. Vedi R.E. Goodin, *Manipulatory Politics*, Yale University Press, New Haven 1980, p. 9 (dove si esamina l'idea della "interferenza sconosciuta"). Gli esempi dati nel testo indicano che la manipolazione può essere un prodotto del consenso e addirittura di un invito.

12. T.M. Wilkinson, "Nudging and manipulation", in *Political Studies*, 61, 2013, p. 345.

13. Vedi J. Elster, *Ulisse e le sirene*, tr. it. il Mulino, Bologna 1983.

14. Vedi R.E. Goodin, *op. cit.*, pp. 7-12.

15. Per risultati pertinenti, vedi G. Felsen et al., "Decisional enhancement and autonomy: Public attitudes toward overt and covert nudges", in *Judgment and Decision Making*, 8, 2012, p. 203.

16. S. Conly, *op. cit.*, pp. 149-172, presenta una serie di casi in cui, dal suo punto di vista, la coercizione è giustificata, fra le altre cose anche per motivi di benessere. Se la manipolazione passa i test pertinenti (che coinvolgono analisi costi-benefici e costi-efficacia, *ibidem*, pp. 150-152), sarebbe giustificata sulla base di motivi analoghi. La stessa Conly avanza qualche osservazione cautelativa a proposito della manipolazione, evidenziando la sua relazione incerta con l'autonomia (*ibidem*, pp. 30-31).

3

IL DIRITTO A NON ESSERE MANIPOLATI

Nel 1890, Louis Brandeis e Samuel Warren hanno proposto un nuovo diritto: il diritto alla privacy.[1] Osservavano che la società moderna crea una grande quantità di nuove minacce al diritto delle persone di essere lasciate ai fatti loro e sottolineavano l'importanza dell'intimità, della possibilità di mantenere spazi privati. Notando che il diritto alla privacy ha molte sfaccettature, sostenevano che le società libere debbono riconoscere un diritto ampio e generale: "Cambiamenti politici, sociali ed economici comportano il riconoscimento di nuovi diritti […]. Ora il diritto alla vita significa il diritto a godere della vita – il diritto a essere lasciati soli".

Il loro ragionamento era e rimane immensamente influente. Parlava di ciò che preoccupava quell'epoca, ma ancor più parla di ciò che preoccupa la nostra epoca. Le minacce alla privacy non sono le stesse, ovviamente, ma l'idea generale di un diritto alla privacy ha contribuito a orientare la discussione e il modo di intendere le cose per ben oltre un secolo.

Ricordiamo questa affermazione della Casa Bianca nel 2024: "Gli americani non devono essere vittime di pratiche online fonte di confusione, manipolative o ingannevoli".[2] Deve esistere un diritto a non essere manipolati? Quel diritto va visto come qualcosa di paragonabile al diritto alla privacy nel 1890? Penso di sì, ma (come si usa dire) è una cosa complicata.

Consideriamo da questo punto di vista alcuni esperimenti istruttivi condotti da Jamie Luguri e Lior Strahilevitz,[3] che erano interessati ai cosiddetti pattern oscuri e ai loro effetti sui consumatori. In uno dei loro esperimenti, hanno testato quale effetto potesse semplificare per le persone

la scelta di un programma costoso di protezione dai furti di identità. Nella condizione di controllo, le persone potevano scegliere il programma, se lo volevano, ma dovevano sceglierlo facendo un clic su un pulsante. In una delle condizioni di trattamento, per gli utenti accettare il programma era più semplice, perché era la scelta predefinita: non dovevano fare clic su alcun pulsante. Era più difficile rifiutarlo, perché era necessario fare clic su "altre opzioni" e poi vedere come fare. L'accettazione del programma, inoltre, era evidenziata come "consigliata". Era uno schema non molto prepotente, ma produceva uno spostamento notevole a favore del programma: lo accettava l'11,3% delle persone nella condizione di controllo, ma il 25,8% nella condizione di trattamento. Va notato che l'intervento moderatamente manipolativo aveva un effetto molto maggiore sui consumatori meno istruiti che su quelli più istruiti.

Luguri e Strahilevitz hanno testato anche l'effetto di vari altri pattern oscuri, progettati per promuovere il programma costoso. Hanno appurato che, se le informazioni sul prezzo erano nascoste in un testo scritto in piccolo e in grigio, aumentava il numero delle persone che lo sceglievano. Hanno scoperto anche che, se l'accettazione era la scelta predefinita, e se le persone dovevano fare uno sforzo in più per rifiutare il programma, il tasso di accettazione aumentava nettamente. Hanno trovato che aumentava significativamente anche se veniva posta una domanda trabocchetto, che rendeva difficile ai consumatori comprendere che cosa stessero scegliendo o rifiutando. È degno di nota che un aumento del costo del programma non aveva alcun effetto sulla vulnerabilità delle persone alla manipolazione online. Si tratta, quindi, di dati empirici che indicano chiaramente come le pratiche manipolative possano avere un grande effetto su chi è messo di fronte a una scelta.

PARLIAMO DI DIRITTI

La maggior parte dei sistemi legali riconosce un diritto a non essere ingannati. Se per "inganno" intendiamo l'indurre intenzionalmente le persone ad avere false credenze, possiamo vedere che, almeno nell'ambito commerciale, l'inganno è in generale contro la legge. Se commercializzo una pillola come "una cura garantita per la calvizie" e la pillola non fa proprio nulla per curare la calvizie, probabilmente ho agito in violazione

della legge. Se vi vendo un nuovo tipo di aspirina e sostengo che previene il cancro, mentre non fa nulla del genere, probabilmente posso essere punito. È vero che l'inganno non viene sempre perseguito. Se vi induco ingannevolmente a credere che il cambiamento climatico non è reale, non mi succede nulla di male. Un motivo è che la libertà di parola richiede un po' di spazio per gli errori, e addirittura per falsità deliberate. Torneremo sull'argomento.

Ora dovremmo poter essere d'accordo che una categoria di azioni, più o meno identificabile, va considerata manipolazione; che quelle azioni non rispettano la capacità di agire autonomamente delle persone; che quelle azioni spesso per quel motivo sono inaccettabili; che quelle azioni sovente conducono le persone in direzioni che peggiorano la loro vita, forse perché chi manipola è una persona malvagia o che fa solo il proprio interesse, magari perché chi manipola è in una posizione di svantaggio epistemico rispetto a chi deve compiere una scelta. I manipolatori cercano di sedurre le persone perché facciano ciò che loro (i manipolatori) desiderano, e spesso hanno successo.

In queste circostanze, c'è un'argomentazione forte a favore di un diritto a non essere manipolati. Dovremmo considerare quel diritto radicato in un principio di autonomia decisionale, sostenuto per ragioni diverse da kantiani e welfaristi e rivolto contro le istituzioni, sia private che pubbliche. Le aziende non devono manipolare i loro clienti. I governi non devono manipolare i loro cittadini.

Ho segnalato casi in cui il diritto a non essere manipolati può essere revocato. Un terrorista, o una persona che commette abusi domestici, perde quel diritto. I diritti morali, di qualsiasi tipo siano, però, non sono assoluti. Il diritto alla libertà di parola non è assoluto, e lo stesso vale per il diritto alla libertà di culto. Un diritto a non essere manipolati è particolarmente importante nell'epoca moderna, in cui la capacità di manipolare le persone, in particolare online, è di gran lunga maggiore di quanto sia mai stata.

MERCATI E INFORMAZIONE

In questa luce, dovremmo essere molto felici di avere una norma sociale contraria alla manipolazione, e potremmo volerla rafforzare, magari per

proteggerci contro commercianti o politici che mirano solo al proprio interesse.[4]

Friedrich Hayek, il maggiore critico della pianificazione in stile socialista, ha esplorato la relazione fra libero mercato e conoscenza. Alcuni dei contributi più importanti di Hayek alla riflessione sociale sono contenuti nel suo grande (e breve) saggio del 1945, "L'uso della conoscenza nella società".[5] In quel saggio, Hayek sosteneva che il grande vantaggio dei prezzi è che aggregano sia le informazioni sia i gusti di molte persone, incorporando una quantità di materiale di gran lunga superiore a quella che potrebbe assemblare qualsiasi organo centrale, pianificatore, gruppo o comitato. Hayek sottolineava la natura non condivisa dell'informazione, i "frammenti dispersi di conoscenza, incompleta e spesso contraddittoria, che gli individui possiedono separatamente". Quella conoscenza sicuramente comprende fatti sui prodotti, ma anche preferenze e gusti, e di tutte queste cose un mercato che funzioni bene tiene conto. Hayek sottolineava soprattutto la "conoscenza, molto importante ma non organizzata, che non può essere considerata scientifica, perché non è conoscenza di leggi generali: la conoscenza delle circostanze particolari di tempo e di luogo".

Per Hayek, sul piano economico la domanda fondamentale è come incorporare quella conoscenza non organizzata e dispersa. È un problema che nessuna persona, nessun gruppo particolare potrebbe risolvere. Chi è responsabile di una pianificazione centralizzata non può avere accesso a tutte le conoscenze di cui dispongono le diverse persone. Presa nel suo complesso, la conoscenza che possiedono quelle persone è di gran lunga maggiore di quella che possiedono anche i più saggi e i più selezionati fra i gruppi o gli esperti. Il punto centrale di Hayek è che la soluzione migliore spesso arriva dal sistema dei prezzi. La sua idea è che, in un sistema in cui la conoscenza dei fatti pertinenti è dispersa fra molte persone, i prezzi si comportano come un dispositivo di coordinamento e segnalazione incredibilmente conciso e accurato. Incorporano quella conoscenza dispersa e, in un certo senso, la rendono anche pubblica, perché il prezzo in sé funge da segnale per tutti.

Ancora meglio, il sistema dei prezzi ha una qualità meravigliosamente automatica, particolarmente nella sua capacità di rispondere rapidamente ai cambiamenti. Se nuove informazioni mostrano che un prodotto (un televisore, un'auto, un telefono cellulare, un orologio) non sempre funziona come dovrebbe, la domanda di quel prodotto diminuirà rapidamente, e lo

stesso vale per il prezzo. Quando un bene di consumo diventa improvvisamente più scarso, i suoi utenti devono rispondere a quel fatto. Secondo Hayek, il mercato funziona notevolmente bene nel suo complesso, non perché qualsiasi partecipante possa vedere tutte le sue caratteristiche, ma perché le informazioni pertinenti vengono comunicate a tutti attraverso i prezzi.

Hayek quindi sostiene che "è più di una metafora descrivere il sistema dei prezzi come una sorta di macchina per la registrazione dei cambiamenti, o come un sistema di telecomunicazione che consente ai singoli produttori di sorvegliare solo i movimenti di pochi indicatori".[6] Hayek descrive questo processo come "prodigioso" e aggiunge che ha scelto quella parola di proposito, "per scuotere il lettore dalla distratta sufficienza con cui spesso diamo per scontato il funzionamento di questo meccanismo". Secondo Hayek, il sistema dei prezzi è un dispositivo straordinario per catturare l'intelligenza collettiva, in parte perché raccoglie quello che ciascuno sa, in parte perché impone gli incentivi giusti.

Per i nostri scopi, questo è l'elemento centrale: tutto questo dipende da un sistema che è relativamente libero da inganno e manipolazione. Se i consumatori invece ne sono vittima, la meraviglia non sarà più tanto meravigliosa. E anche se inganno e manipolazione sono occasionali anziché molto diffusi, molte persone possono esserne danneggiate. Forse le pressioni del mercato li impediranno, o almeno ne ridurranno l'incidenza. Sotto ipotesi ottimistiche, è giusto; però possiamo immaginare facilmente situazioni in cui le pressioni del mercato favoriscono inganno e manipolazione, anziché il contrario. Se in particolare la manipolazione è remunerativa, allora le aziende che non la praticano si troveranno in condizioni di svantaggio competitivo. Il che sembra proprio vero in molti mercati.

VAGHEZZA ED ECCESSO DI POTERE

Anche così, potremmo essere tentati di pensare che la manipolazione ricade in una categoria di azioni opportunamente promosse o scoraggiate da norme sociali, ma in genere non accompagnate da leggi o regole formali. Dal lato positivo, possiamo parlare di gentilezza, considerazione,[7] civiltà, grazia e generosità. Nessuna legge richiede queste cose. Sul fronte negativo,

possiamo parlare di scortesia, inciviltà, malagrazia e meschinità. Nessuna legge punisce la scortesia, la mancanza di considerazione, la malagrazia e la meschinità. Ma perché?

Il primo motivo è la *vaghezza*. Molte norme, fra cui quelle che proteggono i diritti morali, sono rivolte a condotte che non sono, e non possono essere, definite con precisione sufficiente per la legge. Questo problema solleva due preoccupazioni. La prima è che le persone non avranno il giusto preavviso su ciò che possono e non possono fare; non saranno in grado di pianificare. La seconda è che le autorità che devono applicare la legge avranno margini indebitamente ampi di scelta arbitraria. Le persone potrebbero scoprire di partecipare a una lotteria: potrebbero essere prescelte senza alcuna ragione ovvia. Oppure potrebbero scoprire di vivere in un incubo: potrebbero essere selezionate solo per il colore della pelle o per il loro genere.

La vaghezza è un problema reale per un diritto a non essere manipolati. Abbiamo visto che nel linguaggio comune, e anche nelle elaborazioni dei filosofi più attenti, la manipolazione non può essere facilmente il fondamento di norme di legge o del Codice penale. Sono contento delle mie definizioni, ma non posso dire che la definizione (1) nel capitolo 1, o la definizione (2), dovrebbero far parte delle norme di legge.

Il secondo motivo è l'*eccesso di potere*. Molte categorie di cattive condotte comprendono una grande varietà di azioni, alcune delle quali non sono tanto cattive da meritare una punizione, e alcune delle quali potrebbero non essere necessariamente cattive in tutte le circostanze. Le persone inoltre hanno un diritto alla cattiveria, purché non sia eccessiva. Questo è uno dei motivi per cui la mancanza di gentilezza, l'inciviltà e la sconsideratezza non sono contro la legge. Come abbiamo visto, questo sicuramente vale anche per la manipolazione. Un mondo senza manipolazione non sarebbe divertente. Chi vorrebbe non essere mai sedotto?

Potremmo pensare che termini come "frode" e "violenza" abbiano problemi simili, e l'idea non sarebbe del tutto sbagliata. Tuttavia, esistono tradizioni di lunga data che specificano e concretizzano quei termini, riducendo di molto le preoccupazioni relative a vaghezza ed eccesso di potere. Lo stesso però non si può dire della manipolazione.

Forse vi state chiedendo quale sia la differenza fra vaghezza ed eccesso di potere. Supponiamo che una legge proibisca qualsiasi discorso non

protetto dal Primo Emendamento alla Costituzione degli Stati Uniti. Quella legge non sarebbe un caso di eccesso di potere. Per definizione, non permette alcun discorso che il Primo Emendamento non protegge! Comunque, una legge del genere sarebbe di una vaghezza inaccettabile. Il motivo è che il Primo Emendamento è straordinariamente complicato e per molti aspetti poco chiaro; non darebbe alle persone il giusto preavviso.

MIRARE ALLA MANIPOLAZIONE OLTRAGGIOSA

C'è comunque ancora molto che si possa fare, partendo dal concentrarci sui casi peggiori, le pratiche più oltraggiose, definite come quelle che è più difficile giustificare e che è più probabile provochino un danno reale. Abbiamo visto che in alcuni di questi casi il comportamento manipolativo è una forma di furto e, in quei casi, deve essere proibito per quel motivo. (Potete mettere a confronto dire bugie, che non è proibito, con la calunnia, che lo è.) La cosa migliore da fare è focalizzarsi su pratiche specifiche di comportamento manipolativo, quelle che sono chiaramente dannose e molto difficili da difendere con qualsiasi motivazione.

Si sta prestando moltissima attenzione a una varietà di pratiche online che vanno sotto il nome di *pattern oscuri*[8] o schemi ingannevoli. Il termine, inteso a catturare inganno e manipolazione, comprende i casi seguenti.

1. **Scarsità** ("solo 15 rimanenti!").
2. **Dimostrazione sociale** ("Taylor Swift ne va matta!" o "Entra a far parte dei milioni di clienti soddisfatti!").
3. **Messaggio di attività** (informazioni sul livello di attività in un sito, per esempio visualizzazioni o visite).
4. **Conteggio alla rovescia** ("ancora un'ora soltanto!").
5. *Drip pricing* **ovvero incremento graduale del prezzo** (sono 200 dollari – ma poi ci sono oneri aggiuntivi al momento del checkout).
6. **Costi nascosti** (i costi vengono presentati dopo, o sono oscurati).
7. **Informazioni o termini nascosti** (per esempio, pagamenti mensili automatici, maxi-rate finali, rinuncia alla privacy).
8. **Abbonamento nascosto** (un pagamento ricorrente, non indicato chiaramente).

9. *Sneak into basket* (l’aggiunta di prodotti o servizi al carrello senza chiedere il consenso).
10. **Ostruzione** (rendere difficile rifiutare o vedere determinati termini contrattuali).
11. *Roach motel* (un solo clic per abbonarsi, una telefonata o più per disdire).
12. **Azione forzata** (rendere necessario che le persone dichiarino esplicitamente di rifiutare di aderire a qualche programma molto costoso, prima di completare l’acquisto).
13. *Confirmshaming* (“Non ho a cuore la mia privacy” o “Non mi interessa di proteggere la mia salute”).
14. **Asfissiare/bombardare** (“acquista subito, altrimenti perderai l’occasione”, indicato ripetutamente, magari con messaggi di testo).

Alcune di queste pratiche sono di gran lunga peggiori di altre. Prendiamo, per esempio, una pratica della campagna per Donald Trump nel marzo 2020, che corrisponde al caso (7) nell’elenco precedente. Le persone che facevano una donazione alla campagna venivano indirizzate a un modulo online che, in mezzo a molte altre clausole scritte in caratteri più piccoli, conteneva una casella già spuntata con la dicitura “Make this a monthly recurring donation” (“Trasforma questa donazione in una donazione mensile”).[9] Nel giugno dello stesso anno, la campagna utilizzava una strategia simile, denominata “The money bomb”: un’impostazione predefinita per cui le persone si impegnavano a un’ulteriore donazione per il compleanno di Trump.[10] In settembre, la campagna ha trasformato la dicitura “a monthly recurring donation”, con la casella già spuntata, in “a weekly recurring donation” (“una donazione settimanale”).[11]

Con il prosieguo della campagna, la casella “weekly recurring donation” è stata resa meno evidente e spostata al di sotto di altro testo, combinando quindi i pattern (6) e (7) dell’elenco precedente.[12] Queste strategie hanno funzionato. Hanno creato un picco di donazioni, fra cui un gran numero di pagamenti non voluti, che, in molti casi, superavano i limiti delle carte di credito, costringendo le persone a pagare alle banche interessi sullo scoperto, tanto che qualcuno ha dovuto rinunciare alla carta di credito per evitare i pagamenti ricorrenti. Secondo un calcolo, l’uso di *dark defaults* (impostazioni predefinite oscure), come sono state chiamate, ha aumentato

il ricavo di 42 milioni di dollari. Cosa degna di nota, i donatori non hanno dato segno di avere imparato molto dall'esperienza.[13]

Quando le caselle già spuntate, una forma del pattern (7) nell'elenco precedente, vanno considerate una manipolazione e una forma di furto? Per rispondere, dobbiamo prima prendere in considerazione altre due domande: (1) "le caselle già spuntate sono del tutto trasparenti e visibili?" e (2) "fanno l'interesse di tutte o della maggior parte delle persone che ne sopportano le conseguenze?". Se la risposta ad ambedue le domande è positiva, non c'è manipolazione e non abbiamo un problema. Se la risposta a entrambe è negativa, c'è manipolazione, e abbiamo un problema serio. In effetti, dovremmo trovare manipolazione se la risposta alla prima domanda è "no", anche se la risposta alla seconda è "sì". Il motivo è che le persone non sono messe nella condizione di effettuare una scelta deliberativa. Se la risposta alla prima domanda è "sì", ma quella alla seconda è "no", non abbiamo manipolazione (anche se abbiamo un altro tipo di problema).

È importante vedere che, di per sé, le caselle già spuntate non sono particolarmente manipolative, e possono essere un'ottima idea. Un telefono cellulare o un laptop hanno l'equivalente delle caselle già spuntate: sono cose come le impostazioni predefinite, che sono una benedizione. Se alle persone viene detto esplicitamente che un'azienda che vende un servizio online (per esempio Netflix, Apple o Amazon) rinnoverà automaticamente il loro abbonamento, non è affatto chiaro che la legge debba intervenire. Certo, è importante se la maggior parte delle persone, avendo la possibilità di scegliere, fosse a favore di un rinnovo automatico.

I problemi più gravi sorgono quando alle persone *non viene indicato chiaramente che si stanno impegnando a rispettare determinati termini contrattuali* – e quando è molto improbabile che i termini pertinenti siano quello che le persone avrebbero scelto, se avessero saputo di poter scegliere. Come regola generale, dovrebbe essere proibito fare addebiti alle persone in accordo con termini a cui non hanno dato un consenso attivo, certamente nei casi in cui non avrebbero dato il consenso, se fosse stato loro richiesto.[14] Questo è uno dei più oscuri fra i pattern oscuri, ed è una forma di furto. Deve essere proibito. Anche il *drip pricing* è una forma di manipolazione e quindi di furto; a sua volta dovrebbe essere proibito.

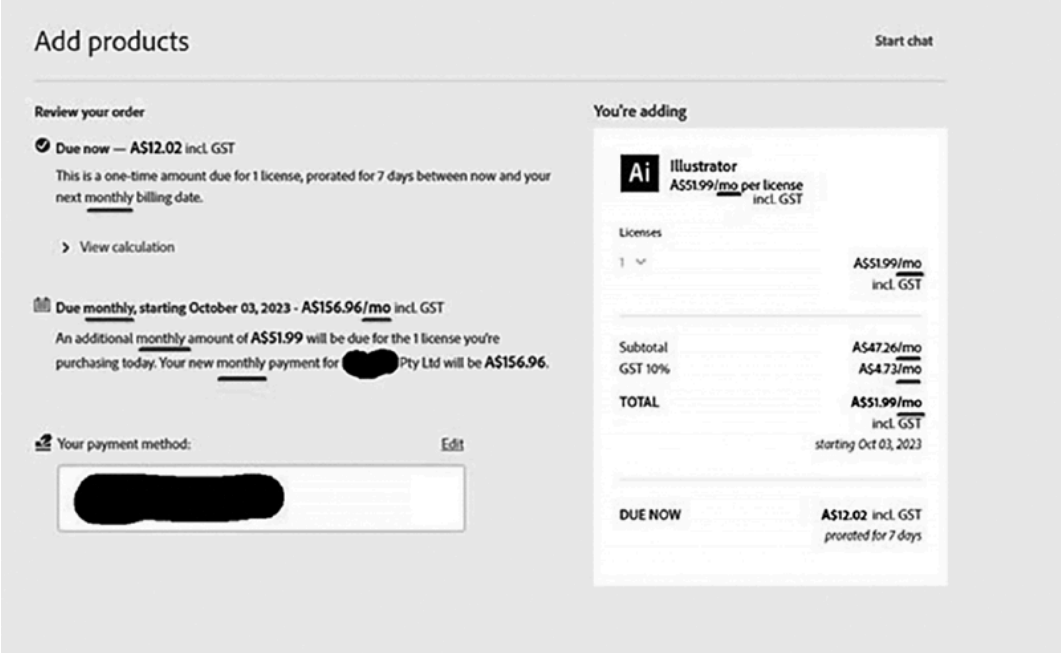

Figura 3.1 Abbonamento a prodotto Adobe.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

Guardate la figura 3.1. Le condizioni di abbonamento sono tutt'altro che chiare, e deve essere stato fatto intenzionalmente.

Nel caso della figura 3.1, l'utente si impegna a un abbonamento annuale, con fatturazione mensile.[15] Se non riuscite a trovare questa informazione nell'immagine, non è colpa vostra. L'unico intervallo temporale visualizzato è la scadenza di pagamento mensile. Non è affatto chiaro che l'utente stia stipulando un abbonamento annuale.

Per proteggersi contro la manipolazione, il principio di fondo deve essere un principio di autonomia decisionale, il che significa che oneri e costi nascosti rientrano in una categoria proibita. Gli organismi di regolamentazione devono prendere di mira saldamente quegli oneri e quei costi. In un'azione rappresentativa nel 2012, il Dipartimento dei Trasporti degli Stati Uniti ha imposto che le compagnie aeree e le agenzie di viaggio online modificassero le loro interfacce web incorporando direttamente nelle offerte pubblicizzate tutte le tasse sui biglietti. L'obiettivo è consentire ai consumatori di compiere scelte informate. I dati disponibili indicano che questo intervento informato sul piano comportamentale, pensato per

eliminare i prezzi nascosti, è stato molto efficace e pertanto ha fatto risparmiare ai consumatori una grande quantità di denaro.[16]

Nel 2024, il Dipartimento dei Trasporti americano ha emanato una norma finale che impone alle compagnie aeree e alle agenzie di viaggio di dichiarare esplicitamente gli addebiti per il bagaglio in stiva, il bagaglio in cabina, le variazioni e le cancellazioni, quando i clienti stanno ancora considerando la possibilità di un acquisto.[17] Sempre nel 2024, la Federal Communications Commission ha imposto ai provider Internet di visualizzare *broadband nutrition labels* ("etichette nutrizionali" per la banda larga) standardizzate.[18] Queste etichette sono pensate per consentire ai clienti di mettere a confronto le offerte dei vari provider punto per punto, e comprendono dettagli standardizzati per i prezzi, le quantità di dati e le velocità in banda larga. Nello stesso anno, il Consumer Financial Protection Bureau (CFPB) ha proposto una norma che limiti gli interessi di scoperto addebitati dagli istituti finanziari con più di 10 miliardi di dollari di asset a una quantità "in linea con i loro costi o in accordo con un benchmark stabilito".[19] La norma proposta dal CFPB includeva un insieme di valori di benchmark, con il massimo fissato a 14 dollari.[20]

UN NUOVO DIRITTO STABILITO

Nel 2024, la Casa Bianca ha annunciato un attacco di tutto il governo ai pattern oscuri di vario tipo, con quella che ha chiamato iniziativa "Time Is Money" ("Il tempo è denaro").[21] Con un accento esplicito sulla manipolazione, questa iniziativa si può vedere come l'incarnazione di un diritto a non essere manipolati. Il governo ha annunciato:

> Gli americani non devono essere bersaglio di pratiche online fonte di confusione, manipolative o ingannevoli.
>
> – Se volete capire che cosa dovete fare per ottenere un bene o un servizio, ciò che è richiesto deve essere chiaro e trasparente.
> – Non dovete essere soggetti a oneri nascosti o a requisiti oscurati grazie a un linguaggio che genera confusione e a testi in caratteri piccoli.

Fermiamoci un attimo. È vero, un semplice annuncio non può creare un diritto di valore legale, ma in queste parole è contenuto il nucleo di un diritto a non essere manipolati.

Fra le altre cose, e coerentemente con l'analisi condotta qui, l'iniziativa attacca termini e oneri occulti o nascosti. Inoltre, prende come bersaglio ciò che descrive come *doom loops*, spirali negative per cui le persone sono "costrette a entrare in un labirinto di opzioni di menu e di registrazioni automatizzate, in cui perdono tempo e non riescono a ottenere il supporto di cui hanno bisogno". In risposta, il governo propone di "imporre alle aziende sotto la sua giurisdizione di consentire ai clienti di parlare con un essere umano con la pressione di un solo pulsante".[22]

Prendendo di mira la manipolazione, il governo nota che le aziende "spesso inducono i consumatori a pagare per abbonamenti – su qualsiasi cosa, dall'iscrizione alla palestra ai quotidiani, ai cosmetici – che non vogliono più o che nemmeno hanno sottoscritto". In risposta, caldeggia una regola che "richieda alle aziende di rendere possibile disdire un abbonamento o un servizio con la stessa facilità con cui lo si può sottoscrivere". Un principio fondamentale è chiaro: "Ottenere rimborsi e risarcimenti deve essere facile come lo è stato l'acquisto, senza scartoffie inutilmente complicate".

Ed ecco un altro principio: "Se si vuole parlare con un essere umano, si deve avere la possibilità di parlare con un essere umano in orari comodi e senza attese interminabili". E un altro: "Non si deve essere assoggettati a oneri nascosti o a requisiti che sono resi oscuri mediante un linguaggio che genera confusione e scritte in piccolo".[23]

Alcuni pattern oscuri, invece, anche se possono essere giudicati manipolativi, non si possono facilmente considerare una forma di furto, e, come regola generale, non dovrebbero essere proibiti dalla legge. Consideriamo tre esempi.

1. Supponiamo che alle persone venga chiesto ripetutamente di fare una certa cosa: "Non vuoi acquistare questo libro?". Le richieste frequenti, il cosiddetto *nagging* ("assillo"), sono irritanti e non si può escludere la possibilità che qualcuno finisca per scocciarsi. Esiste una "spiegazione basata sulla pressione" della manipolazione, in base alla quale alcuni tipi di *nagging* possono essere manipolativi.[24] Un sollecito però non è la cosa peggiore di questo mondo e, se le persone alla fine acconsentono ad acquistare il prodotto o accettano un'offerta, forse è perché, riflettendoci, hanno deciso che è nel loro interesse farlo. Una richiesta frequente può essere un rimedio per un'attenzione limitata e per

l’inerzia, e potremmo anche dare per scontato che molte persone, o quasi tutte, finiscano per compiere una scelta ragionevole. Certo, alcuni casi di *nagging*, che vadano considerati manipolativi o meno, possono essere tanto invadenti, e perciò tanto irritanti, che dovrebbero essere messi fuori legge.

2. Supponiamo che ai consumatori venga detto (e che sia vero) che la disponibilità di un certo prodotto è limitata, o che si utilizzi un conteggio alla rovescia per dare un senso di urgenza (“mancano solo due ore!”). Non è facile vedere queste pratiche, chiamate anche *scarsità*, come un furto, ed è difficile giustificare un intervento legale. Se però le affermazioni che vengono fatte sono false – se le scorte non sono ridotte in modo significativo o se l’offerta non scade tra poche ore – allora siamo di fronte a un caso di frode.
3. Gli usi della *dimostrazione sociale* possono essere manipolativi in qualche caso, ma è difficile capire perché dovrebbero essere proibiti, perlomeno se sono veridici. Se una celebrità apprezza realmente un prodotto, non c’è niente di sbagliato nell’informare le persone di quel fatto. E se i clienti soddisfatti sono molti, non è scorretto rivelare la loro soddisfazione. Certo, gli usi falsi della dimostrazione sociale non sarebbero puramente manipolativi; sarebbero ingannevoli.

Nel pensare se mettere al bando la manipolazione, la cosa migliore che possiamo fare è prendere a esempio i casi più eclatanti, partire da quelli e definire delle categorie a cui si possa fare un riferimento stretto in una regolamentazione. Ricordiamo che le menzogne, come tali, non violano la legge; la violano la calunnia, così come lo spergiuro e la pubblicità falsa. Quando le menzogne sono proibite, è perché causano danni gravi; in qualche caso, sono una forma di furto. Minano alla base o annullano la capacità delle persone di prendere decisioni in autonomia. Tolgono qualcosa ad alcune persone a beneficio di altre. In qualche caso, le pratiche manipolative hanno lo stesso carattere, ed è allora, e per quello, che la legge deve prenderle a bersaglio.

Nel contesto dei prodotti finanziari per i consumatori finali, varie forme di manipolazione sono un problema diffuso. In effetti, la manipolazione può essere vista come una motivazione definitoria per alcune iniziative di regolamentazione. Il Dodd-Frank Wall Street Reform and Consumer Protection Act afferma che il Consumer Financial Protection Bureau (CFPB)

deve garantire che “i mercati per i prodotti e servizi finanziari per i consumatori finali siano corretti, trasparenti e competitivi”.[25] Richiede che si presti attenzione non solo ad azioni e pratiche “scorrette e ingannevoli”, ma anche a quelle “abusive”, il che può essere inteso come un riferimento alle forme peggiori di manipolazione. Nel tenere sotto controllo i mercati pertinenti, il CFPB deve considerare “la comprensione da parte dei consumatori dei rischi di un tipo di prodotto o servizio finanziario per il consumatore finale”,[26] formulazione che si può pensare rispecchi una preoccupazione per la manipolazione.

Nell’implementazione di questi requisiti, il CFPB ha adottato come motto “Know before you owe” (“Informati prima di assumerti oneri”) e i suoi vari sforzi per garantire scelte informate possono essere visti come un attacco alla manipolazione come l’ho definita qui.[27] Nei mercati per i consumatori finali, ovviamente, un problema è la complessità, che può ostacolare la comprensione, e che può essere manipolativa a buon diritto; rivedete le definizioni (1) e (8) nel capitolo 1. Anche altre pratiche ricadono nella categoria generale della manipolazione; pensiamo ai *teaser rates* (i “tassi introduttivi”) e ai vari incentivi che sfiorano l’inganno, ma che sfruttano ampiamente i punti deboli della gente, e che non rispettano la capacità di scelta riflessiva e deliberativa delle persone.

Un contratto per la concessione di una carta di credito, semplice e breve, come quello predisposto dal CFPB, può essere considerato una risposta diretta al rischio di manipolazione[28] – e come un tentativo di garantire che il Sistema 2 abbia le redini saldamente in pugno. Le proposte di proibire o limitare i tassi introduttivi possono essere viste in termini simili.[29] In casi di questo tipo, c’è ampio spazio per considerare il problema della manipolazione nel decidere come regolamentare al meglio i prodotti finanziari. È importante capire che, in quei casi, il governo regola pratiche commerciali, non la pubblicità, e che la sua preoccupazione reale sono pratiche che non rispettano la scelta riflessiva e deliberativa da parte dei consumatori. Le iniziative del CFPB possono essere viste come tentativi di combattere la manipolazione.

REGOLAMENTARE LA LIBERTÀ D'ESPRESSIONE

Secondo la dottrina stabilita, il governo può prendere provvedimenti contro le minacce;[30] può anche regolamentare forme di espressione commerciale false o fuorvianti.[31] Può regolamentare anche la manipolazione? È chiaro che il governo può favorire certe forme di espressione:[32] può favorire forme di espressione che sono verosimilmente manipolative?

Come cartina di tornasole, prendiamo gli sforzi della Food and Drug Administration (FDA) per imporre la presenza sui pacchetti di sigarette di immagini con avvertimenti per la salute. Nel 2011, la FDA ha stabilito una norma che impone una serie di questi avvertimenti (vedi, per esempio, la figura 3.2).

Figura 3.2 Immagini con avvisi di nocività del fumo previste dalla FDA.

Fonte: "New graphic warning labels designed to reduce the deadly effects of smoking", Obama White House Archives, 21 giugno 2011, https://obamawhitehouse.archives.gov/blog/2011/06/21/new-graphic-warning-labels-designed-reduce-deadly-effects-smoking-0.

Questi avvertimenti sono manipolativi in modo non accettabile? Una corte d'appello federale ha invalidato la norma in base al Primo Emendamento, concludendo che la FDA non aveva evidenze sufficienti per giustificare la forma di espressione imposta.[33] Con il suo verdetto, la corte non sottolineava la natura presumibilmente manipolativa degli avvisi minacciosi. L'opinione della corte inferiore però faceva esattamente quello.[34]

Per la corte è risultato rilevante che "le richieste di immagini di avvertimento della norma *non* sono il tipo di affermazioni puramente fattuali e non controverse che sono" oggetto di scrutinio legale meno critico.[35] Aggiungeva, plausibilmente, che "è ampiamente chiaro, dall'esame di quelle immagini, che la risposta emotiva che erano progettate per indurre è calcolata in modo da spingere chi le guarda a smettere di fumare, o a non iniziare mai: un obiettivo del tutto diverso dalla disseminazione di informazioni puramente fattuali e non controverse".[36] La corte concludeva che quando il governo promuove forme di espressione che non riguardano ciò che è "puramente fattuale e non controverso", deve soddisfare un maggior onere di giustificazione. Questa, quindi, è una difficoltà per gli sforzi del governo di imporre una sorta di forma di espressione manipolativa alle aziende private.

L'idea centrale qui non ha un sostegno chiaro nelle decisioni della Corte Suprema, ma ha una certa attrattiva: il Primo Emendamento impone particolari barriere agli sforzi del governo di imporre forme di espressione che non rispettino le capacità deliberative o riflessive, ma che coinvolgono e cercano di attivare il Sistema 1. Da questo punto di vista, non esiste una norma assoluta che proibisca l'imposizione di forme di espressione manipolative di quel tipo (purché non siano false o ingannevoli), ma, se vuole imporre un tale obbligo, il governo deve avere una forte giustificazione.

Questa analisi solleva molti interrogativi diversi. Le immagini come avviso di rischio sono da considerarsi manipolative? Certamente sono pensate per creare una risposta viscerale (e fanno esattamente questo). L'interrogativo però è se rispettino la capacità di scelta riflessiva e deliberativa delle persone. Qui c'è una domanda empirica: che cosa comprendono, esattamente, le persone vedendo quegli avvertimenti? Supponiamo che, per gran parte della popolazione, la comprensione dei rischi sia effettivamente migliorata: in tal caso, questo è un buon argomento per sostenere che non è coinvolta manipolazione. Supponiamo però che la comprensione non migliori: gli avvertimenti possono comunque essere giustificati?

Potremmo immaginare due giustificazioni. La prima è welfarista: le immagini con avvisi di rischio salveranno un numero significativo di vite, mentre informazioni puramente fattuali avranno un effetto di gran lunga più debole. Se è così, allora quelle immagini hanno una giustificazione

sufficiente. La seconda giustificazione fa appello all'autonomia: i fumatori, e i potenziali fumatori, non valutano in misura sufficiente i rischi del fumo per la loro salute, e le immagini di avvertimento possono promuovere una sorta di riduzione delle distorsioni sistematiche del pensiero che le informazioni statistiche non riescono a determinare.[37] A questo punto, si può aggiungere che le norme definite dal governo non vengono imposte su una tabula rasa. Ricordiamo che gli sforzi per promuovere il fumo comportano un grado elevato di manipolazione, con la rappresentazione di fumatori felici, sani e attraenti, e forse il governo può rispondere legittimamente. Considerato il numero delle vite a rischio e l'evidenza empirica relativa, questi tipi di giustificazione sembrano sufficienti, nel contesto particolare del fumo.

Figura 3.3 Etichette della FDA di avvertimento sui rischi del fumo, 2019.

Fonte: Food and Drug Administration. I vari documenti sono consultabili all'indirizzo www.fda.gov/tobacco-products/labeling-and-warningstatements-tobacco-products/cigarette-labeling-and-healthwarning-requirements.

La FDA nel 2019 ha prodotto una nuova serie di immagini di avvertimento per le confezioni di sigarette, fra cui quelle riprodotte nella figura 3.3, che sono un po' meno crude di quelle presentate nella figura 3.2.

La FDA ha affermato che robuste evidenze empiriche facevano pensare che avvertimenti di questo tipo informassero i consumatori dei rischi pertinenti, e quindi producessero scelte più informate.[38] La corte d'appello ha confermato le norme che richiedevano quegli avvertimenti, sostenendo che erano "sia fattuali sia non controversi".[39] Le richieste molto sensate della FDA, e la decisione molto sensata della corte, dimostrano che le immagini di avvertimento non sono necessariamente manipolative in modo non accettabile; possono rispettare la capacità di scelta riflessiva e deliberativa delle persone.

---

1. S.D. Warren, L.D. Brandeis, "The right to privacy", in *Harvard Law Review*, 4, 1890, p. 194.
2. Vedi "Biden-Harris Administration Launches New Effort to Crack Down on Everyday Headaches and Hassles That Waste Americans' Time and Money", 12 agosto 2024, https://www.presidency.ucsb.edu/documents/fact-sheet-biden-harris-administration-launches-new-effort-crack-down-everyday-headaches.
3. Vedi J. Luguri, L. Strahilevitz, "Shining a light on dark patterns", in *Journal of Legal Analysis*, 13, 2021, p. 43.
4. Vedi R.E. Goodin, *Manipulatory Politics*, Yale University Press, New Haven 1980.
5. F.A. von Hayek, "L'uso della conoscenza nella società", in *Competizione e conoscenza*, Rubbettino, Soveria Mannelli 2017, pp. 57-72 (i passi citati sono alle pp. 58 e 60). Eccellenti presentazioni del pensiero di Hayek sono P.J. Boettke, *F.A. Hayek: Economics, Political Economy and Social Philosophy*, Palgrave Macmillan, London 2018; B. Caldwell, *Hayek's Challenge: An Intellectual Biography of F.A. Hayek*, University of Chicago Press, Chicago 2004.
6. F.A. von Hayek, "L'uso della conoscenza nella società", cit., pp. 67, 68.
7. Vedi E. Ullmann-Margalit, *Normal Rationality: Decisions and Social Orders*, a cura di A. Margalit, C.R. Sunstein, Oxford University Press, Oxford 2017.
8. Vedi H. Brignull, *Deceptive Paterns*, Testimonium, Eastbourne 2023; A. Mathur et al., "Dark patterns at scale", in *Proceedings of the acm Human-Computer Interaction*, 3, novembre 2019, https://doi.org/10.48550/arXiv.1907.07032.
9. S. Goldmacher, "How Trump steered supporters into unwitting donations", in *New York Times* (aggiornato 7 agosto 2021), https://www.nytimes.com/2021/04/03/us/politics/trump-donations.html; vedi anche N. Posner et al., "Dark defaults: How choice architecture steers campaign donations", manoscritto non pubblicato, 29 ottobre 2022, https://papers.ssrn.com/sol3/paters.cfm?abstract_id=4258478.
10. S. Goldmacher, *op. cit.*
11. *Ibidem*.
12. *Ibidem*.
13. Vedi N. Posner et al., *op. cit.*

14. Sto andando per grandi linee. Per approfondire alcune delle complessità, vedi M.J. Radin, *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law*, Princeton University Press, Princeton 2012.

15. Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

16. Vedi S. Bradley, N.E. Feldman, "Hidden baggage: Behavioral responses to changes in airline ticket tax disclosure", in *American Economic Journal: Economic Policy*, 12, 2020, p. 58.

17. "Biden-Harris Administration Announces Final Rule to Protect Consumers from Surprise Airline Junk Fees", U.S. Department of Transportation, 24 aprile 2024, https://www.transportation.gov/briefing-room/biden-harris-administration-announces-final-rule-protect-consumers-surprise-airline.

18. Vedi B. Fung, "FCC rolls out mandatory 'nutrition labels' for Internet providers' plans", CNN, 10 aprile 2024, https://edition.cnn.com/2024/04/10/tech/fcc-mandatory-nutrition-labels-internet-providers/index.html; "Broadband consumer labels", Federal Communications Commission, 7 maggio 2024, https://www.fcc.gov/broadbandlabels; "Consumer fact sheet: Broadband consumer label required for most internet service providers", Federal Communications Commission, https://docs.fcc.gov/public/attachments/DOC-401799A1.pdf. [Le *broadband nutrition labels* hanno questo nome perché sono modellate sulle etichette nutrizionali obbligatorie sulle confezioni di generi alimentari. NdT]

19. "CFPB proposes rule to close bank overdraft loophole that costs americans billions each year in junk fees", Consumer Financial Protection Bureau, 17 gennaio 2024, https://www.consumerfinance.gov/about-us/newsroom/cfpb-proposes-rule-to-close-bank-overdraft-loophole-that-costs-americans-billions-each-year-in-junk-fees.

20. *Ibidem*.

21. Vedi *supra*, nota 2.

22. Vedi https://www.presidency.ucsb.edu/documents/fact-sheet-biden-harris-administration-launches-new-effort-crack-down-everyday-headaches.

23. *Ibidem*.

24. Vedi R. Noggle, "Pressure, trickery, and a unified account of manipulation", in *American Philosophical Quarterly*, 57, 2020, p. 245.

25. United States Code, Title 12, 5511.

26. United States Code, Title 12, 5512.

27. Vedi "Know before you owe: Credit cards", Consumer Financial Protection Bureau, https://www.consumerfinance.gov/data-research/credit-card-data/know-you-owe-credit-cards.

28. *Ibidem*.

29. O. Bar-Gill, R. Bubb, "Credit card pricing: The CARD Act and beyond", in *Cornell Law Review*, 97, 2012, p. 967.

30. *Watts v. United States*, U.S. 294, 1969, p. 705.

31. *Virginia State Pharmacy Board v. Virginia Citizens Consumer Council*, U.S. 425, 2976, p. 748.

32. *Johanns v. Livestock Marketing Association*, U.S. 5544, 2005, p. 550; B. Nunes, "The future of government-mandated health warnings", in *University of Pennsylvania Law Review*, 163, 2014, p. 177.

33. *R.J. Reynolds Tobacco Co. v. FDA*, 696 F. 3de 1205, D.C. Cir., 2012.

34. *R.J. Reynolds Tobacco Co. v. FDA*, 823 F. Supp. 2d 36, 2011.

35. *Ibidem*, p. 45 (corsivo mio).

36. *Ibidem*.

37. *Ibidem*.

38. Vedi la motivazione all'indirizzo https://www.federalregister.gov/documents/2020/03/18/2020-05223/tobacco-products-required-warnings-for-cigarette-packages-and-advertisements.

39. Vedi *R.J. Reynolds v. FDA*, 2024, all’indirizzo https://www.ca5.uscourts.gov/opinions/pub/23/23-40076-CV0.pdf.

PARTE SECONDA

APPLICAZIONI

4

DEEPFAKE: MANIPOLATI DALLA TECNOLOGIA

Riuscite a capire quali delle immagini nella figura 4.1 sono reali, e quali sono deepfake? (Risposta: sono entrambe deepfake.)

Come quasi sicuramente sapete già, i deepfake sono prodotti di tecniche basate sull'intelligenza artificiale, volte a creare immagini o video in apparenza reali, in cui si possono vedere persone che fanno o dicono qualcosa che non hanno mai fatto o detto.[1] Produrli è facile. Se caricate un po' di immagini in un sito, potete generare un gran numero di deepfake, in cui la persona in questione appare magra, grassa, orribile, bella, elegante o nuda. Se si può trovare un'immagine di una persona, oggi esiste la tecnologia in grado di creare un deepfake che faccia apparire quella persona come un'atleta professionista o un tossicomane, o qualcuno che odia le persone di colore, disprezza il suo paese, si dà al taccheggio o danza selvaggiamente al ritmo di musiche terribili – qualsiasi cosa. La pornografia deepfake è onnipresente.[2] I deepfake possono essere utilizzati facilmente per screditare il vicino, un nemico o il candidato a una carica pubblica che non vi piace; possono essere utilizzati anche per mettere una persona in una luce eccezionalmente favorevole.

Spesso i deepfake sono descritti come "media manipolati", il che significa non che sono manipolativi, ma che è stata messa in atto qualche forma di manipolazione tecnologica. Il problema della manipolazione, come lo intendo qui, è comunque centrale per i deepfake. Manipolano le persone. Se si mostra un famoso cantante pop che usa cocaina, si trae in inganno il Sistema 1 (e magari anche il Sistema 2). Vedere per credere, come si dice. Se vedete una rappresentazione realistica di una cosa che non

è mai successa, dovrete impegnarvi un bel po’ per eliminare la sensazione, da qualche parte della vostra mente, che di fatto sia realmente avvenuta. I deepfake inducono le persone a credere qualche cosa senza rispettare le loro capacità deliberative.

Figura 4.1 Deepfake.

Fonti: “26 deepfakes that will freak you out”, CNET, 2020, https://www.cnet.com/pictures/26-deepfakes-that-will-freak-you-out; S. Thompson, “A.I. is getting better fast. Can you tell what’s real now?”, in *The New York Times*, 2024, https://www.nytimes.com/interactive/2024/06/24/technology/ai-deepfake-facebookmidjourney-quiz.html.

Ovviamente, è corretto e giusto descrivere i deepfake come una forma di inganno. Sono manipolativi, ma sono anche una menzogna.

BUGIARDI

I sistemi politici e legali devono permettere le menzogne? Che tipo di menzogne? Possiamo affrontare il problema attraverso la lente della legge costituzionale degli Stati Uniti, che offre una chiara prospettiva.

Solo nel 2012 la Corte Suprema degli Stati Uniti ha chiarito che affermazioni false e menzogne, in quanto tali, ricadono nell'ambito del Primo Emendamento, che protegge "la libertà di parola". In *United States v. Alvarez*,[3] una Corte estremamente divisa ha sostenuto che il Primo Emendamento proibisce di perseguire penalmente una persona che abbia sostenuto falsamente di avere ricevuto la Medaglia d'onore del Congresso.

Il caso riguardava Xavier Alvarez, un bugiardo cronico che aveva falsamente sostenuto, fra le altre cose, di essere un veterano del Vietnam, di essere stato un agente di polizia, di essere sposato con un'attrice messicana e di essere stato un giocatore di hockey professionista nei Detroit Red Wings. È finito nei guai con la legge quando era membro del Three Valley Water District Board, un ente pubblico con sede a Claremont, in California. Aveva detto:

> Sono stato un marine per 25 anni, ora in pensione. Sono andato in pensione dai Marines nel 2001. Nel 1987, mi è stata conferita la Medaglia d'onore del Congresso. Sono stato ferito molte volte dalla stessa persona.[4]

Nessuna di quelle affermazioni era vera! Una parte era addirittura illegale. L'affermazione di avere ricevuto la Medaglia d'onore violava lo Stolen Valor Act, che rende un crimine quella particolare menzogna. Ciononostante, la Corte Suprema ha stabilito che la menzogna era protetta dal Primo Emendamento.

Nello spiegare perché quella falsità, grottesca e intenzionale, era protetta, la maggioranza della Corte motivava la decisione con queste parole:

> Permettere al governo di decretare che questa affermazione è un crimine, che sia urlata da un tetto o pronunciata in un sussurro appena udibile, equivarrebbe a riconoscere al governo l'autorità di compilare un elenco di argomenti in merito ai quali le affermazioni false sarebbero punibili. Un tale potere governativo non avrebbe alcun principio limitante ben definito. La nostra tradizione costituzionale va contro l'idea che abbiamo bisogno di un Ministero della Verità di Oceania. Se questa legge venisse sostenuta, ci sarebbe un elenco infinito di argomenti che il governo nazionale o gli Stati potrebbero selezionare.[5]

Fra le altre cose, i giudici d'accordo con questa decisione sottolineavano che punire affermazioni false avrebbe impedito il libero dibattito. Aggiungevano che alternative meno restrittive, fra cui il *counterspeech* ("contro-discorso"), potevano favorire gli interessi legittimi dello Stato (come non sminuire il valore del reale conferimento della Medaglia d'onore del Congresso). Ma quanto potrebbe essere d'aiuto il contro-discorso?

Sarebbe d’aiuto nel caso dei deepfake? Questa è proprio una buona domanda; ci ritorneremo al momento opportuno.

Dovremmo essere d’accordo che al governo non dovrebbe essere data senza alcun limite “l’autorità di compilare un elenco di argomenti in merito ai quali le affermazioni false sarebbero punibili”. Se il Congresso stabilisse che è un crimine fare un’affermazione falsa in merito al cambiamento climatico, o all’evoluzione, o sul calcio o il baseball, o sulla storia americana, tedesca o russa, agirebbe in violazione del Primo Emendamento. Un Ministero della Verità in effetti sarebbe in conflitto con il Primo Emendamento. Un motivo è che esiste una differenza fra una pura falsità e una menzogna; lo stato d’animo del parlante conta. (La maggioranza della Corte Suprema ha trascurato questa differenza: Alvarez ha mentito!) Un altro motivo è che, per molte questioni, è giusto dire che il governo deve fare affidamento sul contro-discorso, nella forma di qualche tipo di refutazione, e non sulla censura o sulla punizione.

MANIPOLAZIONE E VIDEO FALSIFICATI

Con “video falsificati” intendiamo prodotti di tecniche grazie alle quali un video reale viene modificato per far sembrare che le persone facciano o dicano qualcosa di diverso da quello che hanno realmente fatto o detto.[6] Un video falsificato può mostrare una persona che sostiene una causa che invece disapprova, che commette un crimine, che mostra disprezzo per il proprio paese, che agisce in maniera inappropriata quando non ha fatto nulla del genere, o che è ubriaca o non in sé per altre ragioni. In qualche caso, i video falsificati sono molto credibili, nel senso che chi li guarda non pensa che siano stati manipolati. Possono rivelarsi dannosi, fuorvianti, manipolativi o calunniosi.[7] Qui parlerò principalmente di deepfake, ma l’analisi vale più in generale per tutti i video falsificati.

Per orientarci, e per avere un senso di quello che esiste ed è destinato a peggiorare ulteriormente, e molto, prendiamo in considerazione qualche caso ipotetico.

– Jane Jones è un’insegnante di scuola superiore. Un deepfake la mostra in una situazione romantica con uno studente.

– Philip Cross è candidato a una carica pubblica. Un deepfake lo mostra esaltare Hitler e l'Olocausto.
– John Simons ha ottant'anni. In un deepfake lo si vede prendere una "pillola energetica" e poi competere, agguerrito e con buoni risultati, in una partita di basket.
– Un deepfake mostra i Beatles che eseguono brani di Taylor Swift.
– Un deepfake mostra la Monna Lisa che parla come una adolescente.
– Un deepfake mostra un Labrador retriever che balla come Michael Jackson.
– Un deepfake mostra il procuratore generale degli Stati Uniti ubriaco e disorientato.

Si può pensare che, in base alla decisione della Corte Suprema degli Stati Uniti nel caso Alvarez, i deepfake non dovrebbero e non possano essere sottoposti a una regolamentazione semplicemente perché sono deepfake. Sono falsità, quindi frutto della fantasia e, in quanto tali, non sono perseguibili a meno che non causino danno.[8] Lo causano?

Per le persone che vengono usate nei deepfake, il danno assume la forma di danno alla reputazione. Un deepfake può essere favorevole o innocuo, ma può anche essere una forma di diffamazione (o qualcosa che ci assomiglia molto); può esporre le persone al ridicolo o al disprezzo nelle loro comunità. In effetti, può essere devastante. Una persona può essere rappresentata nel modo più orribile che si possa immaginare, che si parli di stupidità, criminalità, corruzione o malvagità (o di qualcosa che ha a che vedere con il sesso). Se un deepfake è diffamante, o il suo contenuto verbale equivale a una calunnia, può essere regolato in molte nazioni, fra cui gli Stati Uniti. Oltre a ferire l'individuo preso di mira, può colpire la comunità, per esempio screditando una persona candidata a una carica pubblica e di conseguenza distorcendo il processo democratico. È un punto che dobbiamo sottolineare. Quando viene diffamato uno di noi, molti di noi possono essere colpiti.

I deepfake in quanto tali, però, non sono necessariamente diffamatori. Come abbiamo visto, possono anche essere positivi: possono far sì che le persone appaiano ammirevoli o meravigliose. Per quel motivo possono essere menzogneri o manipolativi. Per esempio, possono mostrare un politico, un attore o un professore interessato al tema della manipolazione, che giocano a tennis o a golf a livello professionale. Possono presentare una

figura pubblica (poniamo, un candidato a una carica pubblica) che fa qualcosa di ammirevole o di eroico. Anche questo può essere dannoso. Un deepfake può non rispettare la capacità di scelta riflessiva e deliberativa delle persone. È ovviamente una forma di raggiro.

Se le persone non credono che un deepfake sia reale, ovviamente, non ci sarà danno. Se un gruppo rock degli anni Sessanta (due dei cui componenti sono morti) viene rappresentato mentre esegue brani di una musicista contemporanea, abbiamo davanti qualcosa di stravagante, di umoristico o di satirico – tutte cose che devono essere protette secondo il Primo Emendamento.[9] Il rischio di danno sorge se e quando le persone pensano che ciò che vedono sia realmente accaduto. Se una persona viene mostrata falsamente in una situazione romantica o mentre sostiene una posizione politica a cui invece è avversa, di fatto può subire un danno.

Sotto questa luce, consideriamo l'affermazione seguente.

> Il governo può e deve regolare o proibire i deepfake, nel rispetto della protezione della libertà di parola, se (1) non è ragionevolmente ovvio o dichiarato esplicitamente e in modo ben visibile che si tratta di deepfake, e se (2) creano un grave danno di reputazione.

Si noti che (2) si basa sulla legge esistente sulla diffamazione.[10]

Ecco una risposta possibile, ispirata dall'opinione prevalente nel caso Alvarez: la risposta migliore a un deepfake sono contro-discorso e trasparenza, non la censura. Questo vale sia per le organizzazioni private, sia per i funzionari pubblici. L'idea di "danno alla reputazione" è troppo vaga.

Un fornitore di social media, come YouTube o Meta, può concluderne che agli utenti della piattaforma deve essere reso noto che i deepfake sono tali. Può respingere l'idea che i video falsificati debbano essere rimossi dalla piattaforma. Può dedurre che gli utenti devono poter vedere quei video, purché ricevano l'informazione che devono metterli in prospettiva. Come ha scritto la maggioranza della Corte Suprema nel caso Alvarez, "il rimedio per un discorso falso è un discorso vero. Questo è il modo in cui devono andare le cose in una società libera".[11] Alcune piattaforme di social media a volte hanno adottato una impostazione di questo tipo, imponendo la presenza di indicazioni per i media manipolati (e anche rimuovendo tweet quando "è probabile che causino danno").[12]

Per quanto riguarda la legge costituzionale, si può dire qualcosa di simile a proposito di ciò che i governi possono fare. Forse dovrebbe essere

richiesto che scelgano l'approccio che protegge al massimo la libertà di parola. Nel caso Alvarez, per esempio, la maggioranza della Corte Suprema spiegava: "Il governo non ha dimostrato, e non può dimostrare, perché il contro-discorso non sarebbe sufficiente a tutelare i suoi interessi. I fatti di questo caso indicano che la dinamica di libertà di parola, contro-discorso, confutazione possono ovviare alla menzogna".[13] E aggiungeva: "Esiste, tuttavia, almeno un mezzo meno restrittivo della libertà di parola con cui il governo potrebbe probabilmente proteggere l'integrità del sistema di onorificenze militari. Un database creato dal governo potrebbe elencare tutte le persone insignite di una Medaglia d'onore del Congresso. Se esistesse un database del genere, accessibile via Internet, sarebbe facile verificare e denunciare le false affermazioni".[14]

Sarebbe estremamente difficile, però, difendere l'idea che questa sia una risposta sufficiente nei confronti dei deepfake. Supponiamo che una persona venga rappresentata mentre fa qualcosa che non ha mai fatto o sostiene una posizione che invece disapprova. Se così fosse, un database accessibile non avrebbe molte possibilità di eliminare il danno. Forse l'approccio protettivo della libertà di parola sarebbe una divulgazione chiara, che informi le persone, prima che vedano il video, che è un falso. Nel caso Alvarez, non dovrebbe esistere una barriera costituzionale a quel tipo di controllo sui deepfake, almeno nel caso di indicazioni sufficienti di danno, e la stessa analisi vale per i video manipolati.

Quali elementi costituirebbero un'indicazione sufficiente? La diffamazione è una buona risposta, ma non è l'unica. Se le persone finiscono per avere l'impressione che una persona abbia fatto qualcosa che in realtà non ha fatto, può esserci danno per quel preciso motivo. E così è ogniqualvolta un deepfake fa sì che le persone credano che qualcuno ha fatto qualcosa che non ha fatto.

UN DIRITTO DI PROPRIETÀ DELLA PERSONA

Supponiamo di concentrarci sulla manipolazione e l'inganno. Il governo dovrebbe andare oltre la divulgazione? Dovrebbe essere permesso farlo? Controlli più aggressivi possono assumere la forma di multe e diffide, comminate magari da una commissione indipendente. Oppure potrebbero seguire un'impostazione da illecito civile, attraverso cause civili basate

sulla legge contro la diffamazione, e creando una sorta di diritto di proprietà di ciascuno sulla propria persona.

Si può essere tentati di rispondere che, in un senso importante, i deepfake e i video manipolati non sono nulla di nuovo. Sono equivalenti a false affermazioni fattuali. Se un deepfake presenta il procuratore generale ubriaco e intontito, forse non è molto diverso – o non è affatto diverso – da questa affermazione, apparentemente credibile: "Ho visto il procuratore generale ubriaco e intontito". Il contenuto proposizionale dei deepfake, nella forma di affermazioni di questo genere, non è regolamentato né proibito (a meno che non esista una motivazione indipendente – per esempio, potrebbe essere diffamatorio o osceno). Perché i deepfake dovrebbero essere trattati in un modo diverso?

Una risposta ragionevole è che i deepfake (e i video manipolati) hanno un tipo peculiare di autenticità; sono più credibili di rappresentazioni puramente verbali. In un certo senso, si auto-autentificano. Per questo sono manipolativi e anche ingannevoli. La mente umana non li liquida facilmente e, se anche lo fa, c'è sempre una sua parte che rimane convinta. Pensiamo al "bias della verità",[15] la tendenza umana fondamentale a credere a quello che ci viene detto. Tendiamo a dare per scontato che gli altri dicano la verità, perché di solito è così. Supponiamo, per esempio, che arriviate in una città che non avete mai visitato e chiediate indicazioni a qualche estraneo. Probabilmente non vi passerà per la testa che cerchi di mandarvi fuori strada.

Se vedete un deepfake di un procuratore generale ubriaco, avrete qualche difficoltà a togliervi quell'immagine dalla mente. L'immagine vi manipola. Per queste ragioni, è plausibile dire che i deepfake (e i video manipolati) sono correttamente oggetti di attenzione regolativa, anche se le affermazioni che incarnano il loro contenuto proposizionale non lo sono. Come abbiamo visto, poi, le false affermazioni fattuali, gravemente dannose nei confronti del processo politico, possono essere perseguibili anche se non sono diffamanti.

Per le istituzioni private, un approccio aggressivo sarebbe del tutto ragionevole. Per i governi, dovrebbe esistere un mandato di precisa divulgazione per deepfake e video manipolati; in caso di danno palese, potrebbero anche voler andare oltre.

CHE COSA FARE

Queste conclusioni hanno implicazioni forti per le reti televisive e per piattaforme di social media come Meta, YouTube e X. Si possono immaginare restrizioni che proteggano, separatamente o congiuntamente, processi democratici e vite individuali. Nel 2019, Facebook ha risposto a riflessioni di questo tipo generale con l'annuncio che avrebbe bandito i deepfake.[16] È stato un passo importante nella direzione giusta, ma era sufficiente?

Come sottolineava Facebook, i media possono essere manipolati per ragioni innocue, per esempio per rendere le immagini meglio definite e l'audio più chiaro. Alcune forme di manipolazione sono chiaramente intese come battute umoristiche, satire, parodie o affermazioni politiche – come, per esempio, quando una rockstar o un politico vengono raffigurati come giganti o dotati di superpoteri. Questi non sono casi che richiedano una regolamentazione da parte del governo o di Facebook. In base alle sue linee politiche del 2019, Facebook dichiarava che sarebbero stati rimossi "media manipolativi fuorvianti" solo se fossero state soddisfatte due condizioni.

- "Sono stati modificati o sintetizzati – al di là degli aggiustamenti a fini di chiarezza o qualità – in modi che non sono evidenti a una persona media e probabilmente indurrebbero qualcuno, in modo fuorviante, a pensare che la persona rappresentata nel video abbia pronunciato parole che in realtà non ha mai pronunciato."[17]
- "Sono il prodotto di intelligenza artificiale o apprendimento automatico che fondono, sostituiscono o sovrappongono contenuto a un video, facendolo apparire autentico."[18]

In un certo senso, la prima condizione è vicina a quello che ho proposto qui. La persona media, per ipotesi, deve essere ragionevole. Non è prevista fra le condizioni quella di un danno grave (e forse non dovrebbe esistere una condizione del genere, quando si tratta di norme emanate da un'organizzazione privata, e non da un governo). Le due condizioni sono state formulate in modo corrispondente alla preoccupazione di Facebook: l'uso di tecnologie nuove o emergenti per fuorviare la persona media e indurla a pensare che qualcuno abbia detto qualcosa che non ha mai detto.

L’annuncio di Facebook nel 2019 chiariva inoltre che, anche se un video non fosse stato rimosso in base alle nuove norme, potevano essere messe in atto altre salvaguardie.[19] Se, per esempio, un video avesse contenuto violenza esplicita o scene di nudo, sarebbe stato rimosso.

Questo approccio va visto come un progresso importante, e in futuro altre piattaforme di social media faranno bene a tenerlo in considerazione. L’approccio di YouTube è stato a volte più, a volte meno aggressivo; a un certo punto, proibiva senza mezzi termini “contenuti video che siano stati manipolati tecnicamente (al di là di spezzoni presi fuori dal contesto) per costruire eventi in cui è grave il rischio di danni oltraggiosi [*egregious harm*]”.[20] Si noti che YouTube poneva come condizione danni oltraggiosi, cosa che Facebook non aveva fatto. L’espressione è deplorevolmente vaga, ma è chiaramente intesa a consentire media manipolati che siano divertenti o satirici, o che costituiscano nelle intenzioni qualche forma di commento sociale.

Nell’impostazione di Facebook nel 2019, però, rimanevano due problemi. Il primo era che, anche se era coinvolto un deepfake, la norma non si applicava se il deepfake rappresentava fatti anziché parole. (Ricordate la prima condizione, da cui si desume che il deepfake deve presentare il soggetto del video che pronuncia “parole che in realtà non ha mai pronunciato”.) Supponiamo che venga utilizzata l’intelligenza artificiale per mostrare un candidato a una carica pubblica che lavora con terroristi, malmena un bambino piccolo o usa eroina. Nulla nella norma del 2019 prende in considerazione queste rappresentazioni. Questa era una lacuna grave. I deepfake dovrebbero essere banditi dalle piattaforme di social media se presentano azioni che cadrebbero altrimenti nell’ambito della proibizione del 2019, anche se chi è rappresentato non dice nemmeno una parola.

Il secondo problema è che la proibizione era limitata a prodotti dell’intelligenza artificiale o dell’apprendimento automatico. Perché questa limitazione aveva o avrebbe senso? Supponiamo che i video vengano alterati in altri modi – per esempio, rallentandoli in modo da far sembrare che una persona sia ubriaca o intontita da qualche sostanza, come nel caso di un famigerato video manipolato dell’allora portavoce della Camera dei rappresentanti, Nancy Pelosi. Oppure supponiamo che una serie di video, indirizzati contro un candidato alla carica di governatore, venga prodotta non con l’intelligenza artificiale o l’apprendimento automatico, ma

comunque in un modo tale da entrare in conflitto con la prima condizione, cioè che siano stati modificati o sintetizzati per fare sì che una persona comune pensi che quella candidata abbia pronunciato parole che in realtà non ha mai detto. Quello che conta non è la particolare tecnologia utilizzata per ingannare le persone, ma se si tratti di un caso inaccettabile di inganno o manipolazione. Perlomeno, così è se abbiamo a che fare con attori privati.

Meta, come ora è chiamato Facebook, ha modificato significativamente le sue pratiche nel 2020.[21] L'impostazione del 2020 si applicava in generale a video, immagini e audio. Così spiegava Meta:

> D'ora in poi, rimuoveremo media manipolati fuorvianti se soddisfano questi criteri:
> – Sono stati modificati o sintetizzati – al di là di aggiustamenti a fini di chiarezza o qualità – in modi che non sono evidenti a una persona comune e probabilmente indurrebbero qualcuno, in modo fuorviante, a pensare che la persona rappresentata nel video abbia pronunciato parole che in realtà non ha mai pronunciato. E:
> – Sono il prodotto di intelligenza artificiale o apprendimento automatico che fondono, sostituiscono o sovrappongono contenuto a un video, facendolo apparire autentico.
>
> Questa norma non si estende a contenuti che siano parodia o satira, o a video che siano stati modificati soltanto per omettere o modificare l'ordine delle parole.

Nel 2024, Meta ha cambiato ancora.[22] Ha deciso di non rimuovere i media manipolati, ma semplicemente di etichettarli. Ha anche deciso di applicare le sue norme più in generale, includendo qualsiasi tipo di immagine, video o audio. Nelle parole di Meta:

> Se stabiliamo che immagini, video o audio creati digitalmente creano un rischio particolarmente elevato di ingannare materialmente il pubblico su una questione importante, possiamo aggiungere un'etichetta in bella evidenza, così che le persone abbiano più informazione e contesto. Questa impostazione generale dà alle persone più informazioni sul contenuto, in modo che possano valutarlo meglio, e in modo che abbiano un contesto se vedono lo stesso contenuto altrove.

Questa non è una risposta adeguata! Almeno nei casi di probabile danno, la rimozione è la risposta appropriata. Certo, gli "standard della comunità" di Meta richiedono la rimozione, se un deepfake è inaccettabile per altri motivi – perché, per esempio, coinvolge bullismo, molestie o nudità. I deepfake che rappresentano persone o eventi in una luce fuorviante, però, possono essere buoni candidati alla rimozione, almeno se non sono parodia o satira. Certo, qui ci sono problemi difficili di tracciamento dei confini. La cosa da chiedersi, come indicavamo sopra, è che tipo di danno creano.

La tabella 4.1 mostra, a colpo d'occhio, una serie di pratiche relative al 2024 (le pratiche ovviamente cambiano frequentemente).

È importante sottolineare che, in tutto il mondo, le istituzioni sia private che pubbliche stanno cercando di ridurre il rischio di inganno e manipolazione attraverso deepfake. La stessa intelligenza artificiale può essere di aiuto; in effetti, è essenziale. In Indonesia, per esempio, l'USAID ha contribuito allo sviluppo del Misinformation Early Warning System, uno strumento basato sull'intelligenza artificiale che identifica contenuti online manipolati e stabilisce come vengano diffusi. Questo strumento ha contribuito a identificare una rete di bot, a quanto pare dalla Cina, che utilizzava nomi utente femminili all'apparenza occidentali per promuovere una narrazione positiva sul modo in cui sono trattati gli uiguri nella provincia dello Xinjiang. Di conseguenza, varie piattaforme di social media hanno eliminato quei contenuti. La capacità della tecnologia di manipolare informazioni cresce a un ritmo accelerato, ma lo stesso vale per la capacità della tecnologia di individuare la manipolazione.

Tabella 4.1 Le norme delle piattaforme di social media riguardo ai media manipolati.

	Quale tipo di manipolazione è considerato?	**Quali tipi di media sono considerati?**	**Deve essere presente qualche rischio di danno?**	**Rimedio per la violazione?**
Meta				
	1. "modificato o sintetizzato, al di là di aggiustamenti a fini di chiarezza o qualità, in modi che non sono evidenti a una persona comune" e 2. "il prodotto di **intelligenza artificiale**[1] o apprendimento automatico, incluse tecniche di apprendimento profondo (per	Vecchia norma: solo video Nuova norma: video, immagini e audio	Nessuno	Etichettatura

	esempio, un deepfake tecnico), che fonde, combina, sostituisce e/o sovrimpone contenuti […], creando [media] che appaiono autentici.”			
TikTok				
	Generato dall’intelligenza artificiale o “modificato” in modo da “mostrare scene o persone che appaiono realistiche”.	“Contenuti, inclusi immagini, video o audio”	Nessuno (ma presupposto di danno per le categorie messe al bando)[2]	Nella maggior parte dei casi: **etichettatura** In casi speciali:[2] **rimozione**
X				
	1. “alterato, manipolato o costruito in modo significativo e ingannevole”; o 2. “condiviso in modi ingannevole con contenuti falsi”.	Qualsiasi medium (“Inclusi immagini, video, audio, gif e URL che ospitano contenuti rilevanti”)	“è probabile che abbia come conseguenza diffusa confusione su questioni pubbliche, impatto sulla sicurezza pubblica, o causi danni gravi”	Nella maggior parte dei casi: etichettatura (o avvertimenti alle persone che mettono un “mi piace” o condividono il post, riduzione della visibilità del post, disattivazione dei “mi piace” e dei commenti)
YouTube				
	“manipolato tecnicamente o modificato in modo da fuorviare gli utenti (di solito al di là di spezzoni presi fuori dal contesto)”	“Contenuti”	“può creare un rischio grave di danno rilevante”	Rimozione

Note: Le citazioni sono tratte dalle fonti seguenti. Per Meta: M. Bickert, “Enforcing against manipulated media”, Facebook, 6 gennaio 2020, https://about.fb.com/news/2020/01/enforcing-against-manipulated-media [https://perma.cc/A4SV-878X]; per TikTok: “Misleading and false content”, TikTok (ultimo aggiornamento novembre 2024), https://ads.tiktok.com/help/article/tiktok-ads-policy-misleading-and-false-content#anchor-1-0; per X: Y. Roth, A. Achuthan, “Building rules in public: Our approach to synthetic & manipulated media”, X Blog, 4 febbraio 2020, https://blog.x.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media; per YouTube: “Misinformation policies”, YouTube Help, 2024, https://support.google.com/youtube/answer/10834785?hl=en.

1. È possibile che nuove norme riguardino un insieme più ampio di media manipolati, ma i post sul blog di Meta fanno pensare che saranno limitate ai contenuti generati dall’intelligenza artificiale.
2. I contenuti generati dall’intelligenza artificiale o i media modificati saranno rimossi se mostrano falsamente: (1) contenuti presentati in modo da sembrare che provengano da una fonte autorevole, per esempio una fonte giornalistica di buona reputazione; (2) un evento di crisi, come un conflitto o un disastro naturale; o (3) una figura pubblica che (a) viene umiliata o molestata, o che esibisce comportamenti criminali o antisociali; (b) assume una posizione su una questione politica, un prodotto commerciale, o una questione di importanza pubblica (per esempio, un’elezione); o (c) viene appoggiata o condannata politicamente da un individuo o un gruppo.

IL MONDO NUOVO

Ogni giorno la tecnologia fa passi avanti e, tra non molto tempo, le capacità tecnologiche saranno avanzate in modo inimmaginabile. Non perdiamo di vista, però, i nostri punti di riferimento nel vortice degli eventi quotidiani. La psicologia umana rimane costante, o quasi, e lo stesso vale per i principi di fondo. I deepfake sono un pericolo evidente per l’autonomia dell’individuo, per l’integrità della reputazione, per il commercio, la politica e la democrazia. Sono spesso ingannevoli. Sono spesso manipolativi. Quando causano danni gravi, non devono essere consentiti.

---

1. M.J. Blitz, “Lied, line drawing, and (deep) fake news”, in *Oklahoma Law Review*, 71, 2018, pp. 59, 61.
2. D. Keats Citron, “Sexual privacy”, in *Yale Law Journal*, 128, 2019, pp. 1870, 1921-1922.
3. 567 U.S. 709, 2012.
4. *Ibidem*.
5. *Ibidem*, p. 723 (il “Ministero della Verità di Oceania” è una citazione da *1984* di George Orwell).

6. Vedi, per esempio, BBC News, “Fake Obama created using AI video tool”, YouTube, 19 luglio 2017, www.youtube.com/watch?v=AmUC4m6w1wo [https://perma.cc/89DTY4C6]; Fortune Magazine, “What Is a Deepfake? Video Examples with Nicolas Cage, Jennifer Lawrence, Obama Show Troubling Trend”, YouTube, 11 marzo 2019, https://www.youtube.com/watch?v=AmUC4m6w1wo [https://perma.cc/WW9F-JZ9F]; B. Warner, “Deepfake video of Mark Zuckerberg goes viral on Eve of House A.I. hearing”, in *Fortune*, 12 giugno 2019, https://fortune.com/2019/06/12/deepfake-mark-zuckerberg [https://perma.cc/52PM-2DGE].

7. *Corp. Training Unlimited, Inc. v. Nat’l Broad. Co.*, 868 F. Supp. 501, 507 (E.D.N.Y. 1994) (spiega che quando esaminano “parti audio e video di un programma televisivo” i tribunali devono essere consapevoli della “possibilità che una trascrizione che sembra a prima vista relativamente inoffensiva possa essere in realtà […] estremamente tossica”, e che “un amalgama sapiente di mezze verità e di affermazioni simili a opinioni, abbellito da immagini opportunamente selezionate e da un accompagnamento audio drammatico, può essere devastante, quando presentato in un medium potente come la televisione”).

8. Nel caso Alvarez, il giudice Breyer notava che “pochi statuti, o forse nessuno, proibiscono semplicemente, senza porre limitazioni, di dire menzogne, anche menzogne su un argomento particolare. Invece, in praticamente tutti questi casi limitazioni di contesto, richieste di dimostrazione di danni e simili, restringono lo statuto a un sottoinsieme di menzogne, laddove è più probabile che si verifichi un danno specifico”. 567 U.S. 709, 736, 2012 (giudice Breyer, a favore). Continuava spiegando che “le limitazioni contribuiscono ad assicurare che lo statuto non consenta che la minaccia di perseguibilità o di condanna penale si possa allargare a piacere, scoraggiando o proibendo la menzogna in contesti in cui il danno è improbabile o la necessità di proibizione è minima” (*ibidem*).

9. *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46, 54, 1988.

10. Anche se i deepfake non minacciano di danneggiare le persone raffigurate, possono comunque provocare danni. Pensiamo, per esempio, a un deepfake che rappresenta un candidato a una carica pubblica che fa qualcosa di eroico, quando invece non ha mai fatto nulla del genere.

11. *Alvarez*, 567 U.S., p. 727.

12. “Building rules in public: Our approach to synthetic & manipulated media”, Twitter Blog, 4 febbraio 2020, https://blog.x.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media. Il testo pubblicato sul blog precisa quanto segue.

Alcuni danni specifici che consideriamo:

- minacce alla sicurezza fisica di una persona o di un gruppo;
- rischi di violenza di massa o di diffuse sollevazioni civili;
- minacce alla privacy o alla possibilità di esprimere liberamente se stessi (per una persona o un gruppo) o di partecipare a eventi civici, per esempio: stalking o attenzione non desiderata e ossessiva; contenuti mirati che includano modi di dire, epiteti o materiali che cerchino di ridurre qualcuno al silenzio; soppressione o intimidazione di votanti.

13. *Alvarez*, 567 U.S., p. 726.

14. *Ibidem*, p. 729.

15. I. Skurnik et al., “How warnings about false claims become recommendations”, in *Journal of Consumer Research*, 31, 2005, p. 713.

16. M. Bickert, “Enforcing against manipulated media”, Facebook, 6 gennaio 2020, https://about.fb.com/news/2020/01/enforcing-against-manipulated-media [https://perma.cc/A4SV-878X].

17. *Ibidem*.

18. *Ibidem*.

19. *Ibidem*. Vedi anche G. Rosen, “Helping to protect the 2020 US election”, Facebook, 21 ottobre 2019, https://about.fb.com/news/2019/10/update-on-election-integrity-efforts

[https://perma.cc/Q6DV-YA4Q] (dove vengono annunciate ulteriori salvaguardie per ridurre l'impatto della disinformazione virale sulle elezioni negli Stati Uniti).

20. Le line guida e le norme della Community di YouTube sono in continua evoluzione e possono essere verificate direttamente nella sezione del sito dedicata all'assistenza. La frase citata compariva in "Spam, deceptive practices, & scam policies" (YouTube Help, https://support.google.com/youtube/answer/2801973?hl=en), ma ad oggi le linee guida sono cambiate e tale indicazione è stata eliminata. [NdR]

21. "Enforcing against manipulated media", Facebook, 6 gennaio 2020, https://about.fb.com/news/2020/01/enforcing-against-manipulated-media.

22. "Our approach to labeling AI-generated content and manipulated media", Facebook, 5 aprile 2024, https://about.fb.com/news/2024/04/metas-approach-tolabeling-ai-generated-content-and-manipulated-media.

5

SCARSITÀ E ATTENZIONE: MANIPOLATI DAL PANTANO

Gli illusionisti manipolano le persone. Traggono in inganno l'occhio. Fanno in modo che l'attenzione si concentri su un punto e la distolgono da un altro, che invece è quello in cui succedono effettivamente le cose. Sono dei campioni in questo.

Guardate *The Prestige*, lo stupendo film del 2006 diretto da Christopher Nolan. È la storia della strenua competizione fra due illusionisti: Alfred Borden (interpretato da Christian Bale) e Robert Angier (interpretato da Hugh Jackman). Le cose si guastano del tutto quando Borden provoca accidentalmente la morte della moglie di Angier (interpretata da Piper Perabo), il che porta a una competizione fra i due illusionisti, sostanzialmente fino alla morte. (Ma è proprio così?) Borden e Angier dispongono di molti trucchi: in genere traggono in inganno le persone sfruttando il fatto che l'attenzione è limitata.

La limitazione dell'attenzione deriva a sua volta dalla *scarsità cognitiva*, che ci espone alla possibilità di essere manipolati. Un po' di pantano (*sludge*) peggiora di molto le cose, e gli illusionisti ne introducono di vari tipi. Ma sto correndo troppo.

Chi vende online non è un illusionista, almeno non tecnicamente, ma mette comunque in atto un po' di magia. Vari anni fa, la American Express mi ha fatto un'offerta: potevo ricevere, gratuitamente, l'abbonamento a sei riviste. Sembrava un buon affare. (Riviste gratis!) Delle riviste in offerta non ce n'era neanche una che amassi veramente, ma sono riuscito a trovarne sei che sembravano interessanti, e le ho scelte. Dieci anni dopo, stavo ancora ricevendo quelle riviste. A quell'epoca lavoravo sul tema dell'economia comportamentale, e ho deciso di indagare. Ho scoperto che,

da nove anni e nove mesi, stavo pagando l'abbonamento a riviste che non mi piacevano. Semplicemente non avevo prestato attenzione al fatto che l'"abbonamento gratuito" significava tre mesi gratis, poi il costo dell'abbonamento mi sarebbe stato automaticamente addebitato, a meno che non avessi disdetto.

Chi lo sapeva? Io no. La American Express aveva cercato di manipolarmi, e ci era riuscita.

Più recentemente, ho raccontato questa vicenda a un pubblico di persone d'azienda, fra le quali alcune che collaboravano alla gestione di una rivista settimanale. Una poi è venuta da me e mi ha detto: "Lo sappiamo bene. Se l'abbonamento si rinnova automaticamente, il nostro tasso di rinnovo è straordinariamente elevato, oltre il 90%. Se invece chiediamo alle persone se vogliono rinnovarlo, beh, perdiamo un sacco di soldi".

Mentre scrivevo questo libro, ho avuto la fortuna che mi venisse chiesto di tenere una conferenza in una università ben nota. La conferenza sarebbe stata remunerata – una cifra certo non enorme, ma neanche molto piccola. Con mia sorpresa, l'università mi ha inviato un link a una cosa chiamata Service Portal, e mi ha chiesto di registrarmi come "fornitore". Il portale era così complicato, con una navigazione tanto difficile, che alla fine ho rinunciato a tentare di utilizzarlo. E lo stesso vale per il mio favoloso assistente. Dopo varie richieste online dall'università, e note poco amichevoli (spedite dall'intelligenza artificiale?), non ho visto altra possibilità che offrirmi di tenere la conferenza gratuitamente.

Ora, prendiamo in considerazione i casi seguenti.

1. Un'azienda produttrice di telefoni cellulari commercializza molti dei suoi modelli con un rimborso via posta. I consumatori hanno diritto a un rimborso di 200 dollari. L'azienda sa bene che molti consumatori saranno attratti dal potenziale rimborso – e che non riusciranno mai a spedire per posta i moduli. L'azienda è molto felice di questo fatto.
2. I consumatori possono abbonarsi alle trasmissioni della Amazing Science Fiction Network, una rete televisiva; l'abbonamento dà loro accesso a molti programmi, fra cui anche spettacoli che non vengono trasmessi in televisione. Ottenere un abbonamento è facile: basta sostanzialmente un clic. Recedere dall'abbonamento non è altrettanto facile: bisogna fare una telefonata, restare in attesa per un bel po' di tempo e poi rispondere a una serie di domande (forse poste da alieni

venuti dallo spazio), teoricamente pensate per assicurarsi che le persone vogliano realmente disdire l'abbonamento.

3. Per ottenere dei benefici in base a una legge sull'assistenza sanitaria, le persone devono navigare in un sito web complicato. Spesso non capiscono le domande che vengono loro poste. Per molte, inoltrare una richiesta richiede parecchio tempo. Alcune rinunciano.
4. Per presentare un reclamo per un'automobile difettosa, per legge i consumatori devono compilare lunghi moduli online. I moduli richiedono informazioni dettagliate su dove il veicolo è stato acquistato inizialmente e su come è stato usato. Alcuni consumatori non hanno facilmente accesso a quelle informazioni. Altri temono una violazione della loro privacy. Molti decidono di non compilare i moduli.
5. Ogni volta che visitano certi siti web, le persone sono informate sulle regole approvate da quei siti in tema di privacy e, se vogliono procedere, devono spuntare una casella, per indicare il proprio consenso. A volte il documento sulla privacy, legalmente obbligatorio, è in una lingua che gli utenti non conoscono.

Tutti questi casi comportano *sludge*, una "miscela viscosa" o un "pantano", sotto forma di oneri amministrativi o attriti che rendono difficile, per consumatori, dipendenti, datori di lavoro, studenti, pazienti, clienti, piccole imprese e molti altri, ottenere ciò che vogliono o fare ciò che desiderano. L'imposizione di un pantano può essere una forma di manipolazione. A volte è addirittura magica.

Imponendo un pantano, le organizzazioni pubbliche e private possono generare molta frustrazione e rendere difficile alle persone compiere una scelta riflessiva e deliberativa. Spesso il pantano è invisibile; le persone non vi hanno dato alcun consenso. Il pantano può rendere un'interazione, online o meno, molto diversa da quella che le persone si aspettavano e volevano.

In questo capitolo ho tre obiettivi. Il primo è sviluppare l'idea di pantano, con riferimenti stretti all'economia comportamentale. Come vedremo, il pantano è un modo straordinariamente efficace di manipolare le persone.

Il secondo obiettivo è dare una spiegazione del perché il pantano può essere tanto dannoso per consumatori, studenti, pazienti, clienti, investitori, lavoratori e altri ancora, facendo riferimento agli studi comportamentali su bias del presente, inerzia, attenzione limitata e ottimismo irrealistico. Le

organizzazioni, pubbliche e private, a volte hanno un forte interesse ad aumentare il pantano – e a volte invece un forte interesse a ridurlo.

Il terzo obiettivo è chiedere regolari Sludge Audit, nel settore privato come in quello pubblico. La finalità di questi Sludge Audit è fare chiarezza sull'ordine di grandezza dello *sludge* e garantire trasparenza alle persone che ne sono influenzate, incluse quelle che sono nelle condizioni di ridurlo. Grazie a quella trasparenza, gli Sludge Audit devono motivare una valutazione attenta, da parte della giurisprudenza e della politica, di come ridurre il pantano – e una forma pervasiva di manipolazione, specialmente dei consumatori.

È importante notare che gli Sludge Audit potrebbero essere più o meno formali ed elaborati. Potrebbero essere fortemente quantitativi, con un tentativo di calcolare costi e benefici. Potrebbero essere più qualitativi, orientati a capire che cosa viene richiesto e a chiedere, in un modo più intuitivo, se i livelli di pantano esistenti sono eccessivi. In tutti i casi, dovrebbero essere visti come parte di uno sforzo generale di individuare dati comportamentali al fine di semplificare la giurisprudenza, il governo e la politica.

UN'OCCHIATA

Per un'occhiata preliminare ai potenziali benefici degli Sludge Audit, prendiamo il programma PreCheck della Transportation Security Administration (TSA), pensato per alleggerire le code ai controlli di sicurezza negli aeroporti. Il programma, adottato dagli Stati Uniti nel 2011, seguiva una sorta di Sludge Audit; era basato sulla conclusione che i processi standard di sicurezza sono eccessivamente onerosi per molti viaggiatori. Le persone che aderiscono al programma possono godere del vantaggio di controlli più rapidi e code meno lunghe – un chiaro tentativo di riduzione del pantano. In un anno recente, erano iscritte al programma 20 milioni di persone. Quali sono i benefici?

Ipotizziamo, prudenzialmente, che in media quei 20 milioni di persone utilizzino il programma quattro volte all'anno: parliamo, allora, di 80 milioni di utilizzi. Ipotizziamo anche, plausibilmente, che un utente del programma risparmi venti minuti per viaggio: si tratterebbe allora approssimativamente di 1,6 miliardi di minuti risparmiati all'anno.

Ipotizziamo, infine, che un'ora valga in media 30 dollari: il beneficio del programma della TSA sarebbe dunque di almeno 48 miliardi di dollari all'anno. Per eseguire un'analisi completa degli effetti di benessere del programma, ovviamente, dovremmo conoscere anche i costi. Impostare e gestire il programma richiede risorse significative, ma si può tranquillamente dire che, una volta che il programma è in funzione, i benefici netti sono molto elevati.

Le code ai controlli di sicurezza non sono esattamente manipolative, e un po' di pantano non è una forma di manipolazione, ma l'esempio è comunque illustrativo. Ogni volta che si può far risparmiare una certa quantità di tempo (poniamo, 200.000 ore) a una popolazione numerosa di consumatori, i benefici non saranno proprio banali. Cosa importante, monetizzare i risparmi di tempo è tutt'altro che sufficiente per individuare quei benefici. Alcuni dei benefici della riduzione del pantano sono psicologici e edonici – una riduzione della frustrazione, dell'ansia e magari anche di un senso di stigma o di umiliazione. Inoltre, i tentativi di ridurre il pantano possono migliorare molto l'accesso a beni e servizi, come permessi, licenze, programmi contro la povertà, istruzione, formazione al lavoro e opportunità economiche.

Per consumatori, studenti e lavoratori, iniziative di questo genere possono cambiare la vita. Per le aziende, possono far aumentare le vendite e la reputazione. Per i governi, possono migliorare molto le prestazioni. Possono promuovere sviluppo economico e crescita. In molti contesti, le iniziative per ridurre il pantano avranno benefici notevoli per i componenti più svantaggiati della società; possono essere un motore di opportunità. Le nazioni più varie, che abbiano a cuore la lotta alla povertà e l'istruzione, farebbero bene a condurre Sludge Audit, e a fare della riduzione del pantano una priorità.

PANTANO COME MANIPOLAZIONE

Se intendiamo il pantano come costituito da oneri o attriti amministrativi, il concetto non è affatto misterioso. Il termine deve essere preso come riferimento al tipo di attrito, piccolo o grande, che le persone devono vincere se vogliono andare in una direzione o nell'altra. Per le loro ragioni, egoistiche o altruistiche che siano, le istituzioni private e pubbliche

possono imporre o aumentare il pantano. Nel settore privato, le aziende possono utilizzarlo per aumentare i profitti. Per esempio, le persone possono voler annullare l'abbonamento a una rivista a cui non sono più interessate, ma, per poterlo fare, devono navigare faticosamente attraverso una marea di pantano. Questo è un modo di manipolare le persone. Il pantano online è una sorta di pattern oscuro; è un modo per indirizzare le persone verso esiti che richiedono loro tempo o denaro.

Nel settore pubblico, il pantano può essere accidentale, ma può essere anche il frutto di una scelta deliberata, nel qual caso può essere considerato una forma di manipolazione. Una persona vuole ottenere per i figli qualche facilitazione, per esempio trasporti gratuiti o mensa scolastica gratuita, ma il pantano può scoraggiarla. Può darsi che l'intenzione sia effettivamente quella. Il diritto di votare può essere il più fondamentale di tutti, ma un processo di registrazione pieno di pantano può tenere lontane dalle urne milioni di persone. Può essere anche manipolativo, una forma di raggiro pensata per impedire alle persone di compiere scelte riflessive. Un'iniziativa di riduzione del pantano può essere una "legge sul diritto di voto". Per ottenere un aiuto finanziario per il college, gli studenti negli Stati Uniti devono compilare il Free Application for Federal Student Aid (FAFSA). Fino a poco tempo fa, era un modulo lungo e complicato, e i ragazzi dovevano fornire informazioni di cui potevano non essere a conoscenza (alcune riguardavano la denuncia dei redditi dei genitori). Molti studenti gettavano la spugna, e forse l'obiettivo era proprio questo.

Per chiarezza concettuale, non dobbiamo considerare pantano gli incentivi o disincentivi monetari. Se si dice ai consumatori che devono pagare una certa cifra per ottenere un'assicurazione o che possono ottenere un posto migliore su un aereo pagando una piccola cifra aggiuntiva, questi sono costi, non pantano. Un divieto legale non è pantano: se è proibito fumare nei locali pubblici, non c'è un problema di pantano. Un obbligo legale può essere pantano e può non esserlo, a seconda di ciò che è reso obbligatorio. Se alle persone si dice che devono superare un processo inutilmente complesso per avere un aiuto per la depressione o l'ansia, sicuramente ciò che è richiesto è navigare in mezzo al pantano. Se alle persone viene detto che devono avere una assicurazione sanitaria, vale la stessa conclusione, se ottenere l'assicurazione comporta una grande quantità di attrito.

Da un certo punto di vista, ovviamente, non esiste una distinzione reale fra incentivi monetari e i tipi di costi imposti dal pantano. In fin dei conti, il tempo può essere monetizzato e, se i costi sono di natura emotiva o edonica, anche quelli possono essere monetizzati. Un costo monetario di una certa quantità ridotta è minore del costo di una certa quantità elevata di tempo. Se siamo convinti che i costi sono semplicemente costi, allora il pantano può essere caratterizzato come un tipo di costo, il cui ordine di grandezza può essere quantificato e monetizzato.

L'unica risposta è che le distinzioni qualitative sono utili, ed esiste una distinzione qualitativa importante fra (poniamo) un'imposta o una contravvenzione da una parte e la richiesta di compilare un modulo inutile dall'altra. Nel settore pubblico, il pantano è spesso un metodo indiretto, nascosto e poco sottoposto a controlli per raggiungere particolari obiettivi politici, limitando l'accesso a programmi e benefici; è un metodo che andrebbe messo bene a nudo.

Il pantano è un male se e quando impone attriti *eccessivi*. Deve essere chiaro che gli attriti o gli oneri amministrativi, imposti da istituzioni private o pubbliche, possono essere eccessivi, insufficienti o ottimali. Se a una persona viene chiesto se è sicura di voler cancellare qualche file importante, di rinunciare ai propri diritti legali, di inviare un messaggio di posta incollerito, o di pubblicare qualcosa sui social media, gli oneri risultanti possono essere uno sforzo per impedire errori o mosse avventate. Un periodo obbligatorio di ripensamento per le vendite porta a porta, come quello imposto dalla Federal Trade Commission nel 1972, è un esempio. Così come lo è un periodo di attesa prima di completare l'acquisto di un'arma. Domande del tipo "sei davvero sicuro di volerlo?" possono portare benefici sia agli individui che alla società. Un periodo di attesa può addirittura essere un modo per indebolire l'effetto di un tentativo di manipolazione. Al momento opportuno, spiegherò quali siano le ragioni legittime per il pantano.

PERCHÉ IL PANTANO È IMPORTANTE

Deve essere chiaro che chi impone del pantano può farlo intenzionalmente, ma anche involontariamente. Dal punto di vista morale, l'imposizione intenzionale di pantano può essere particolarmente

deplorevole. Chi impone pantano spesso cerca di promuovere il proprio interesse. Spesso l'imposizione di pantano è una forma di manipolazione, per esempio quando le aziende cercano di sfruttare l'inerzia o il bias del presente per portare le persone ad acquistare prodotti o servizi inutilmente costosi. Quando l'imposizione di pantano è involontaria, non deve essere vista come una forma di manipolazione, ma può comunque essere molto negativa. Può essere il prodotto di scarsa lungimiranza, come, per esempio, quando burocrati o legali impongono oneri amministrativi che si rivelano avere effetti avversi gravi o addirittura devastanti, anche se chi aveva introdotto quegli oneri non aveva intenzione di ottenere quel risultato. Anche se i benefici del pantano sono elevati, i suoi costi possono dimostrarsi soverchianti.

Questi costi possono assumere forme qualitativamente differenti. Possono comportare l'acquisizione di *informazioni* difficili e costose da ottenere. Possono comportare *tempo*, che le persone magari non hanno. Possono essere costi *psicologici*, nel senso che comportano frustrazione, stigma o umiliazione. (Persone povere, anziane o disabili possono sentirsi particolarmente frustrate, stigmatizzate o umiliate.) Quando questi costi sono elevati, può diventare molto difficile navigare fra il pantano o vincerlo. In qualche caso, compilare tutta la modulistica pertinente può essere letteralmente impossibile; semplicemente le persone potrebbero non essere in grado di ottenere tutte le informazioni necessarie per compilare i moduli, o di capire come farlo. Queste cose contribuiscono a spiegare le basse percentuali di adesione a molti programmi statali e federali negli Stati Uniti, e anche le immense difficoltà che le persone spesso incontrano quando cercano di ottenere permessi o licenze di vario tipo.

Dovremmo considerare il pantano anche un ostacolo alla libertà, specialmente nella misura in cui riduce o ostacola la navigabilità. La libertà di scelta è importante ma, se le persone non possono arrivare dove vorrebbero andare, non sono esattamente libere. Il pantano può operare nello stesso modo della coercizione. Crea muri e barriere.

Abbiamo visto che un tipo di pattern oscuro è quello del "roach motel": facile entrare, difficile uscire. Questa è una forma di manipolazione. La figura 5.1 mostra una finestra dal sito della SiriusXM[1] che richiede ai clienti di telefonare (e a volte di rimanere a lungo in attesa) per disdire il loro abbonamento, una difficoltà a cui vanno incontro gli abbonati a molti tipi di prodotti e servizi.

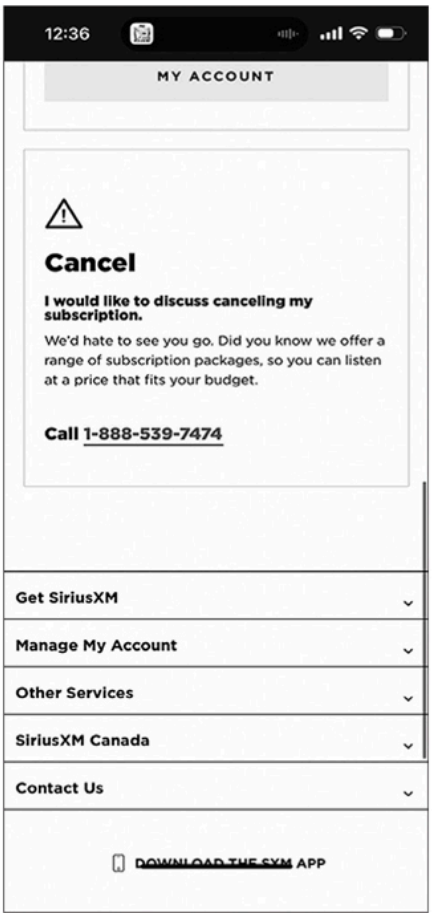

Figura 5.1 Come disdire un abbonamento a SiriusXM.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

Negli Stati Uniti, l'importante iniziativa del 2024 "Time Is Money",[2] che attaccava la manipolazione, era esplicitamente progettata per proibire questo tipo di pratica.

PANTANO E MANIPOLAZIONE IRRIMEDIABILI

Molti bias umani amplificano gli effetti reali del pantano. Per molte persone, l'inerzia è una forza potente, e tendono a procrastinare. Se una persona è soggetta all'inerzia, e continua a procrastinare, potrebbe non completare mai la modulistica necessaria. Il problema è aggravato dal bias del presente. Quando è intenzionale, il pantano sfrutta i bias umani, e per questo motivo è manipolativo.

Per capire perché il pantano è importante, notiamo che per molti di noi il futuro assomiglia a un paese straniero – Futurlandia – che non sappiamo se mai visiteremo. Spesso è forte la tentazione di rimandare le faccende amministrative a un altro giorno. Quel giorno può non arrivare mai, anche se le conseguenze del ritardo sono molto gravi. Quanti impongono pantano spesso sono felicemente consapevoli di questo.

Prendiamo i buoni da spedire per posta, senza dubbio una forma di pantano. Tali buoni danno la possibilità di ottenere qualche tipo di vantaggio, spesso sotto forma di un assegno, ma per utilizzarli le persone devono vincere l'inerzia. Un'illustrazione del rapporto fra bias comportamentali, pantano e manipolazione arriva da uno studio sul comportamento delle persone a proposito del riscatto o meno dei coupon, studio dal titolo memorabile: "Everybody believes in redemption".[3]

Nei vari mercati, i tassi di redemption di solito vanno dal 10 al 40% circa, il che significa che una stragrande maggioranza dei clienti se ne dimentica o semplicemente se ne disinteressa. Data la forza dell'inerzia, questo forse non sorprende particolarmente. Ciò che colpisce di più è il fatto che le persone siano ottimiste in modo non realistico in merito alla probabilità di riscattare prima o poi i coupon. Nello studio citato, le persone pensavano che la probabilità di farlo entro i trenta giorni dal ricevimento fosse dell'80%, mentre il tasso reale di redemption era del 31%. È eccessivo dire che tutti credono nel riscatto – ma la maggior parte delle persone sicuramente ci crede.

Nello stesso studio, i ricercatori hanno provato in tre modi (con gruppi diversi di persone) di ridurre la grande differenza fra i tassi di redemption previsti e quelli reali. In primo luogo, hanno informato i partecipanti, in modo molto chiaro, che in gruppi precedenti, con persone simili, i tassi di redemption erano inferiori a un terzo. In secondo luogo, hanno presentato due chiari promemoria, uno subito dopo l'acquisto e un altro quando la scadenza utile per il riscatto si avvicinava. In terzo luogo, hanno reso più semplice il riscatto eliminando il requisito di stampare e firmare una pagina di certificazione.

Nessuno dei tre interventi ha ridotto l'ottimismo delle persone! In tutte le condizioni, pensavano che la probabilità che inviassero per posta i moduli fosse di circa l'80%. Inoltre, cosa che ha lasciato un po' sorpresi, i primi due interventi non avevano alcun effetto su ciò che le persone facevano effettivamente. Quando sentivano del comportamento di altri gruppi,

evidentemente pensavano: “Beh, quelli sono *altri* gruppi. Che cosa hanno a che fare con noi?”. In altri contesti, i promemoria spesso funzionano perché concentrano l’attenzione delle persone e riducono il potere dell’inerzia; in questo caso, invece, i promemoria si sono dimostrati privi di utilità.

L’unico intervento efficace è stata la semplificazione, che aveva un forte impatto su ciò che le persone facevano effettivamente. Semplificando l’invio del modulo e quindi riducendo il pantano, la semplificazione aumentava significativamente la propensione ad agire delle persone. Il tasso di redemption è aumentato a circa il 54%, il che significa che la disparità fra credenza e comportamento si riduceva alla metà. Qui, dunque, abbiamo una prova concreta dell’effetto potenzialmente significativo della riduzione del pantano nell’aumentare l’accesso delle persone a benefici preziosi. Vale la pena prendere in considerazione, in un contesto come questo, se una riforma legale sia una buona idea.

BIAS COMPORTAMENTALI, MANIPOLAZIONE E *SLUDGE*

Bisogna ricordare che l’inerzia è una forza potente e che, a causa sua, le persone possono non compilare i moduli necessari o navigare altrimenti in mezzo al pantano. Questo è uno dei motivi per cui i tassi di partecipazione spesso sono più bassi quando il consenso deve essere esplicito (*opt-in*) rispetto a quando l’adesione è l’opzione predefinita (*opt-out*). Va ricordato anche che l’inerzia è aggravata dal bias del presente, per cui le persone si concentrano sul breve termine e trascurano il futuro. Supponiamo a questo punto che i consumatori debbano compilare determinati moduli per poter esercitare il diritto a benefici importanti o per evitare penalizzazioni significative. Possono avere l’intenzione di farlo, ma se l’incombenza può essere rimandata, o se è onerosa o difficile, il loro comportamento può non corrispondere alle intenzioni. I costi effettivi si possono dimostrare molto elevati; i costi percepiti possono esserlo ancora di più. Più in generale, il pantano ha un impatto significativo che molte persone non prevedono. Come mostra lo studio sulla redemption, le persone sono ottimiste in modo non realistico in merito alla loro probabilità di vincere l’inerzia. Anche gli specialisti possono rimanere sorpresi da quanto falliscano anche strategie apparentemente promettenti.

Sotto questo profilo, il pantano può essere utilizzato opportunisticamente da commercianti manipolativi per cercare di dare ai consumatori l'impressione di ricevere un'offerta eccellente, ben sapendo che i consumatori non coglieranno l'opportunità. In qualche caso, i funzionari governativi fanno la stessa cosa: sbandierano la disponibilità di un certo beneficio, o mascherano i loro sforzi di ridurne la disponibilità, nel contempo utilizzando il pantano in modo opportunistico. In alternativa, o magari allo stesso tempo, possono rispondere a valori e impegni politici. I funzionari pubblici possono utilizzare il pantano come un dispositivo di razionamento o come un modo per conservare le risorse dei contribuenti. In qualche caso, il pantano ha un effetto dannoso che i funzionari non prevedevano. In particolare, le organizzazioni private e pubbliche possono non comprendere in quale misura il pantano inciderà negativamente su una popolazione che stanno cercando di aiutare. I bias comportamentali possono rendere essenzialmente proibitivi gli oneri amministrativi o imporre una difficoltà grave in termini di navigabilità; le persone, nel settore privato come in quello pubblico, possono non riuscire a prevederlo.

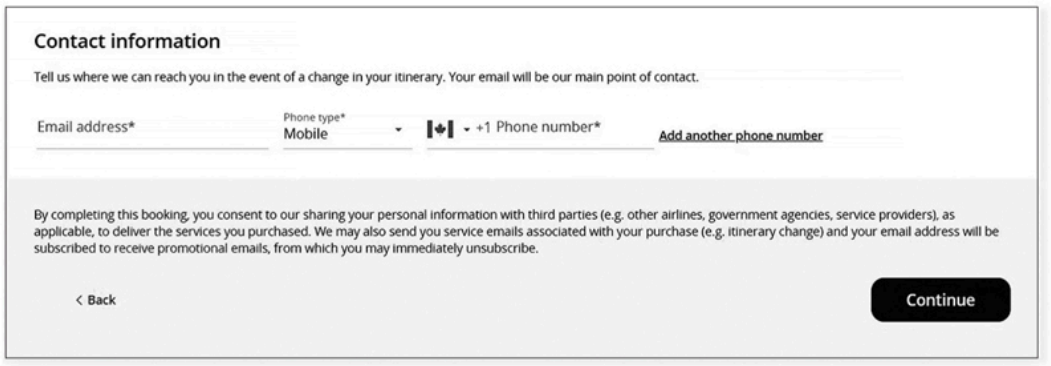

Figura 5.2 E-mail promozionali dell'Air Canada.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-ofshame.

Consideriamo l'impegno necessario per disdire l'iscrizione a e-mail promozionali. Tutti amano il Canada (o dovrebbero – io di certo lo amo), ma a quanto pare l'Air Canada sembra sfrutti il pantano e l'inerzia per catturare un pubblico rendendo automatica e obbligatoria l'iscrizione alle e-mail promozionali quando si acquista un biglietto online (vedi la figura 5.2).

SCARSITÀ

Rispetto ai tassi di conversione delle offerte, l'efficacia della semplificazione getta nuova luce sulla gravità delle conseguenze di un pantano anche apparentemente modesto – sugli effetti dell'architettura delle scelte nel determinare gli esiti. Semplificazione e riduzione degli oneri non si limitano semplicemente a ridurre la frustrazione; possono cambiare la vita delle persone. Un motivo di fondo è che le nostre risorse cognitive sono limitate. Inevitabilmente, siamo in grado di concentrarci solo su un piccolo sottoinsieme delle sfide della vita. Per chi è molto impegnato, povero, disabile o anziano, il problema della scarsità cognitiva è particolarmente grave.[4] Per questo è importante concentrarsi sugli effetti *distributivi* del pantano – sulle persone che hanno più probabilità di esserne colpite. Utilizzato per manipolare le persone, il pantano può incidere negativamente soprattutto su quelle che si trovano alla base della scala sociale ed economica.

In pratica, l'onere del pantano, imposto dalla legge, spesso è sopportato principalmente dalle persone povere, ed è un onere che non possono sostenere facilmente. Un motivo centrale è che, a causa della loro povertà, le persone povere devono concentrarsi su un ventaglio ampio di problemi immediatamente pressanti. Se un'organizzazione privata o pubblica chiede a una persona povera di navigare in un sistema complesso o di compilare un gran numero di moduli, quella persona può rinunciare. Il problema però non si limita solamente ai poveri. Quando ci sono programmi pensati a beneficio delle persone anziane, il pantano può essere particolarmente dannoso, almeno se la popolazione interessata soffre di una ridotta capacità cognitiva. Qualcosa di simile si può dire per gli immigrati o per le persone che sono ostacolate da una barriera linguistica. Per motivi diversi, il problema dell'uguaglianza di genere richiede un'attenzione particolare. Poiché le donne svolgono una quantità molto maggiore di lavoro amministrativo (gestendo l'economia domestica, preparando i pasti, prendendosi cura dei figli), una riduzione significativa del pantano potrebbe aiutarci ad affrontare una fonte onnipresente di disuguaglianza sociale, che ha effetti ramificati su altri ambiti della vita.

GIUSTIFICARE IL PANTANO

Nonostante tutto quello che ho detto fin qui, il pantano può essere funzionale a obiettivi importanti. Può aiutare consumatori, pazienti, studenti e lavoratori. A volte può essere giustificato con motivazioni legate al benessere. Non è per forza manipolativo; per certi versi, può addirittura rafforzare l'autonomia delle persone.

Problemi di autocontrollo

Abbiamo visto che il pantano può essere progettato in modo da contrastare problemi di autocontrollo, avventatezza e impulsività. Può essere un modo per proteggere le persone contro i loro stessi errori. Può dare più forza al pianificatore di fronte a un esecutore non abbastanza deliberativo. Per questo motivo può funzionare come un pungolo, come cura di un problema comportamentale. I pungoli o *nudges* sono definiti come *interventi rispettosi della libertà che orientano le persone in una particolare direzione*. Per esempio, un avviso è un pungolo, e lo stesso vale per un po' di attrito (per altri particolari, vedi i capitoli 7 e 8).

Per decisioni quotidiane, i pungoli si incontrano spesso online, sotto forma di domande che chiedono se "vuoi davvero" spedire una e-mail senza oggetto, annullare un ordine recente o cancellare un file. Questi pungoli possono essere un'idea eccellente, e non devono per forza essere manipolativi. Pungoli analoghi, predisposti da organizzazioni private e pubbliche, possono avere senso per decisioni che cambiano la vita, come matrimoni e divorzi. Periodi di ripensamento possono essere una benedizione. Pensiamo al caso dell'acquisto di armi, dove i periodi di attesa possono salvare vite.

Privacy e sicurezza

Le organizzazioni, private e pubbliche, spesso impongono pantano al fine di ottenere informazioni sulle persone e il loro pregresso – la carriera lavorativa, il reddito, la fedina penale, la valutazione del credito, la storia familiare, i luoghi di residenza. Chi cerca lavoro nel settore pubblico, sicuramente a livelli che riguardano la sicurezza nazionale, deve fornire una grande quantità di informazioni di quel genere. (Ho avuto molte di queste posizioni nella pubblica amministrazione e, anche se il pantano non è molto divertente, è giustificato.) È almeno ragionevole pensare che, se le

organizzazioni, private e pubbliche, devono ricevere alcune o tutte quelle informazioni, ciò deve avvenire con il consenso esplicito delle persone interessate. Se è così, la questione è se chiedere alle persone di affrontare oneri amministrativi o invece violare la loro privacy. Forse non è tanto terribile se i funzionari scelgono la prima opzione. Anche qui, non esiste problema di manipolazione.

Un tempo, ovviamente, i funzionari non avevano una vera possibilità di scelta. Non potevano invadere la privacy perché non ne avevano i mezzi. Sempre più, però, ora le organizzazioni private e pubbliche hanno un accesso indipendente a quelle informazioni, o possono ottenerle con un po' di impegno online. Ma sarebbe desiderabile? Non necessariamente. In qualche caso, bisogna trovare un compromesso fra oneri irritanti da un lato e potenziali invasioni della privacy dall'altro. Pensiamo, per esempio, a quante informazioni dovrebbero poter acquisire le società delle carte di credito prima di offrire una carta a un cliente. Possiamo vedere di buon occhio situazioni in cui quelle società possono scoprire ciò di cui hanno bisogno e poi semplicemente inviano alle persone le loro offerte o addirittura le carte. Che le vediamo di buon occhio o meno dipende in parte da quali informazioni hanno e se possono essere utilizzate male. Se il governo ha, o acquisisce, le informazioni rilevanti, si può pensare che i rischi siano inaccettabili.

Il problema della sicurezza è strettamente collegato. Per creare un conto online, ai consumatori può venire chiesto di fornire (e i consumatori saranno propensi a fornire) informazioni delicate, che riguardano, per esempio, il loro conto bancario o la loro carta di credito. Le persone devono magari rispondere a domande sul loro domicilio, il codice fiscale, o il nome da nubile della madre. La richiesta che le persone rispondano a queste domande può essere una seccatura, ma non è manipolativa, e può essere giustificata come strumento per la difesa da qualche tipo di violazione.

Idealmente, ovviamente, vorremmo avere chiari i benefici e i costi dell'ottenimento delle informazioni pertinenti. Se però costi e benefici sono difficili da specificare, può avere senso avere almeno un'idea che siano desiderabili impegni non troppo onerosi per prevenire gli scenari peggiori in termini di privacy e sicurezza.

Acquisizione di dati utili

Aziende e governi possono cercare di acquisire dati utilizzabili per varie finalità, e questo può andare anche a grande vantaggio del pubblico. Per esempio, le aziende possono cercare di scoprire molte cose sui consumatori, per poter offrire loro un servizio migliore. Oppure i funzionari pubblici vogliono sapere se le persone che ricevono formazione per il lavoro, o qualche tipo di finanziamento per l'istruzione, traggono effettivamente beneficio dai programmi pertinenti. Che cosa fanno con quella formazione o quel finanziamento?

Le persone che ricevono richieste di raccolta di informazioni possono lamentarsi del pantano, ma gli oneri pertinenti possono essere giustificati come mezzi per garantirsi l'acquisizione di conoscenze importanti o addirittura indispensabili. Nell'epoca moderna, l'acquisizione di informazioni può promuovere la responsabilità pubblica e privata. Può fare risparmiare denaro. Può favorire l'innovazione. Può addirittura salvare vite. Non deve per forza manipolare qualcuno.

Idoneità e qualifiche

Quando le organizzazioni, private e pubbliche, impongono pantano, un motivo può essere legato alle restrizioni di idoneità; un altro può essere legato alla gestione di archivi di dati. Chi vuole ottenere un mutuo, privato o pubblico, deve compilare dei moduli. Il motivo principale è verificare, in base alla legge, la sua idoneità. Le persone non devono ricevere benefici governativi a meno che non ne abbiano il diritto, secondo la legge.

È vero che, con la disponibilità crescente di informazioni e con gli algoritmi di apprendimento automatico, le organizzazioni private e pubbliche possono riuscire a scoprire le informazioni pertinenti da sole. Nel settore privato, alcune aziende usano l'idea della "prequalifica", il che significa che hanno abbastanza informazioni da sapere in anticipo che alcune persone già sono idonee a riceve certi beni o servizi. A volte i moduli possono essere parzialmente "precompilati"; di conseguenza, gran parte del lavoro di compilazione può non essere necessaria. Nel campo fiscale, un esempio è l'idea del *return-free filing* ("dichiarazione dei redditi precompilata"), per cui i contribuenti non devono più compilare nulla. Quando i tempi saranno maturi, dovremmo vedere movimenti significativi in questa direzione, ma per il momento siamo solo agli inizi.

La spiegazione più ovvia per il pantano va sotto il nome di “integrità del programma”. Supponiamo che l’Internal Revenue Service decida di inviare lo Earned Income Tax Credit a contribuenti che apparentemente ne abbiano diritto. Se lo può fare a basso costo, e se i contribuenti che apparentemente ne hanno diritto ne hanno effettivamente diritto, non ci sarebbe molto da obiettare. Il problema, ovviamente, è l’avverbio “apparentemente”. È possibile che alcune delle persone che lo ricevono in realtà non ne abbiano diritto. Ogniqualvolta le persone vengono inserite automaticamente in un programma, è possibile che alcune non soddisfino i criteri legali.

Quando le cose stanno così, l’ente regolatore deve scegliere fra (1) un progetto predisposto in modo che alcune persone che ne avrebbero diritto non riceveranno un beneficio e (2) un progetto predisposto in modo che alcune persone che non ne avrebbero diritto ricevano un beneficio. Se la “integrità del programma” si intende riferita al numero degli errori, la scelta fra (1) e (2) può dipendere semplicemente dall’aritmetica: quale gruppo è più numeroso? Se l’inserimento automatico significa che riceveranno il beneficio 500.000 persone che altrimenti non lo avrebbero ricevuto, questa è un’ottima cosa, e se il pantano dovesse impedire anche solo che non ricevano il beneficio 50 persone che non ne hanno diritto, l’inserimento automatico sembrerebbe giustificato.

Se ci mettiamo a giocare con i numeri, però, potremmo immaginare che ci siano ragionevolmente dei disaccordi. Supponiamo che l’inserimento automatico dia benefici a 200.000 persone aventi diritto che altrimenti non li riceverebbero, ma anche a 200.001 persone non aventi diritto che altrimenti non li riceverebbero. Questo sembrerebbe davvero terribile: così tante persone che ricevono del denaro a cui non hanno diritto! Qualcuno però potrebbe pensare che se le 200.001 persone hanno “quasi” diritto – se cioè sono relativamente povere – non è tanto terribile se ricevono qualche aiuto economico. Altre persone potrebbero pensare che il denaro dei contribuiti sia accompagnato da chiare restrizioni, e se viene distribuito violando quelle restrizioni viene commesso un grave torto – il che significa che anche una modestissima violazione dell’integrità del programma, con vantaggio delle persone che non hanno diritto, è inaccettabile.

La versione più estrema di questo punto di vista sarebbe che la concessione di benefici a un numero molto grande di persone che ne hanno diritto non compenserebbe la concessione di benefici a un numero molto piccolo di persone che non ne hanno diritto. Da un punto di vista welfarista,

la versione più estrema è difficile da difendere: la concessione di benefici a dieci persone che ne hanno quasi (ma non esattamente) diritto sembrerebbe un prezzo che valga la pena pagare in cambio della concessione di benefici a un milione di persone che ne hanno effettivamente diritto. Il punto giusto di compromesso, però, non è autoevidente. Anche qui, di nuovo, non siamo necessariamente in presenza di manipolazione.

SLUDGE AUDIT

Ho sottolineato che il problema del pantano va oltre il problema della manipolazione, ma un modo per combattere la manipolazione è condurre regolarmente Sludge Audit. Esploriamoli, dunque, in parte per ridurre la manipolazione, ma più in generale per rendere la vita più facile e migliore.

I governi certamente dovrebbero effettuare questi controlli, e lo stesso vale per una grande varietà di organizzazioni private. Banche, compagnie di assicurazione, ospedali, università, venditori online e editori potrebbero risparmiare tantissimo denaro riducendo il pantano, e potrebbero contemporaneamente migliorare l'esperienza delle persone che interagiscono con loro. Potrebbero addirittura riuscire a cambiare la vita delle persone. Vale la pena evidenziare il caso degli ospedali, dove il pantano può non solo creare frustrazione immensa, ma anche mettere a repentaglio la salute e addirittura costare la vita. Uno Sludge Audit si sovrappone in parte a un Dark Patterns Audit, perché molti pattern oscuri comportano pantano.

Abbiamo visto che gli Sludge Audit possono essere formali o informali. Possono comportare un grande lavoro di quantificazione, o essere più qualitativi. In ogni caso, tre riforme contribuirebbero moltissimo a migliorare la situazione attuale. In primo luogo, le istituzioni dovrebbero intraprendere una "revisione" periodica degli oneri esistenti per vedere se lo "stock" corrente può essere giustificato e per eliminare ciò che si rivelasse obsoleto, inutile o troppo costoso. Questa riforma si baserebbe sui requisiti di revisione esistenti per le norme in generale. In secondo luogo, le istituzioni dovrebbero scegliere il metodo meno oneroso per raggiungere i loro obiettivi. Questo è essenzialmente un requisito di efficacia rispetto ai costi. Se, per esempio, una rendicontazione annuale sarebbe efficace quanto una rendicontazione trimestrale, gli enti dovrebbero scegliere quella

annuale. In terzo luogo, le istituzioni dovrebbero garantire che i benefici del pantano giustifichino i costi.

È vero e importante che un punto di equilibrio fra costi e benefici non è sempre facile da raggiungere per il pantano. La quantificazione può essere problematica, può essere difficile stabilire equivalenti monetari. In tali casi, un approccio brutale sarebbe intendere il requisito della giustificazione costi-benefici non come un tentativo di confrontare costi sociali e benefici sociali, in termini economici, ma invece come qualcosa che comporta una valutazione più intuitiva di *proporzionalità*. Viene imposta una grande quantità di pantano e, se è così, è probabile che serva a finalità significative? Possiamo dire qualcosa dell'ordine di grandezza effettivo dei costi e dei guadagni? Numeri reali aiuterebbero a rendere informate le decisioni e a combattere il pantano, ma dovremmo poter fare molti passi avanti anche senza di essi.

Vale la pena sottolineare il fatto che anche una forma rozza di analisi costi-benefici sarebbe una *spinta all'informazione*. Creerebbe un chiaro incentivo ad avere una contabilità accurata del numero di ore di lavoro rese necessarie dai livelli correnti di pantano, e anche a trasformare quelle ore in equivalenti monetari. Simultaneamente creerebbe per le organizzazioni un incentivo a essere più specifiche, e più quantitative, in merito ai benefici attesi del pantano.

Nei casi più facili, gli Sludge Audit mostrerebbero immediatamente alle organizzazioni, sia pubbliche che private, che il livello esistente di pantano non è nel loro interesse, e che danneggia o manipola le persone. Rispetto al settore pubblico: se si rivela difficile l'accesso per i bambini alla mensa scolastica gratuita, semplicemente perché il pantano è eccessivo, i funzionari possono intraprendere dei passi per ridurlo. Se risulta difficile ottenere aiuti finanziari per gli studenti, perché i moduli prevedono più di 100 domande, rendersi conto di questo fatto potrebbe dare il via a un impegno a ridurre il numero delle domande. Se si stabilisce che famiglie in difficoltà non riescono a ricevere gli aiuti alimentari a cui avrebbero legalmente diritto, la riduzione del pantano, magari attraverso l'uso di servizi online, potrebbe sembrare molto attraente. Se si dà il caso che l'ottenimento di permessi edilizi richiede una quantità immensa di tempo e di fatica, e se quei tempi e quella fatica contribuiscono alla carenza di abitazioni, gli sforzi per ridurre il tempo e l'impegno necessari possono essere accolti da un plauso generale.

Rispetto al settore privato: se ai consumatori risulta difficile fare quello che è necessario per acquistare un prodotto (poniamo, un'automobile), un'azienda può semplificare l'esperienza. Così facendo, potrebbe attirare più clienti e ottenere un ampio ventaglio di benefici di reputazione. I dirigenti dell'azienda potrebbero capire che troppo pantano è un esercizio di manipolazione. Non è proprio una novità che i consumatori abbiano un'esperienza di gran lunga peggiore con un'azienda se è difficile scoprire o comprendere i termini di un accordo, o ottenere una risposta alle loro lamentele. Molte aziende hanno innovato creativamente, nel tentativo di ridurre questi problemi. Potremmo facilmente immaginare una sorta di competizione per diventare un ente pubblico o un'azienda libera da pantano rispetto a tutto ciò che conta per i cittadini o i consumatori. Lo stesso potrebbe valere per lavoratori, investitori e studenti.

All'altro estremo, le organizzazioni pubbliche o private potrebbero sapere, o scoprire, che il pantano è nel loro interesse, e uno Sludge Audit non creerebbe un incentivo a ridurlo. Se il pantano scoraggia l'immigrazione, alcuni funzionari pubblici sarebbero felici di imporre del pantano, magari di aumentarlo. Se il pantano riduce la possibilità di accesso a determinate professioni, i funzionari, favorevoli agli interessi dei professionisti già qualificati, potrebbero non essere particolarmente dispiaciuti. Può far bene ai conti aziendali rendere molto facile sottoscrivere un abbonamento – e pieno di pantano il processo per disdirlo. Test accurati potrebbero mostrare che una strategia del genere è ottimale. Un processo di reclamo che comporti una certa quantità di pantano potrebbe non semplicemente evitare i reclami; potrebbe anche far risparmiare denaro quando i reclami sarebbero giustificati.

Si possono immaginare circostanze in cui il pantano è nell'interesse competitivo delle aziende; è il motivo per cui esiste. I pattern oscuri fanno guadagnare. Se è così, rimane l'interrogativo: questo è un fallimento comportamentale del mercato per cui è appropriata una risposta normativa? La risposta sarà spesso "sì". Ricordate l'iniziativa "Time Is Money"[5] del 2024, che è piena di risposte normative al pantano?

L'aspetto più generale è che, per le istituzioni pubbliche, uno Sludge Audit spesso mostrerà che esistono occasioni significative di migliorare le performance. Ogni anno, per esempio, il governo degli Stati Uniti deve rendere pubblico qualcosa di simile a uno Sludge Audit: lo Information Collection Budget (ICB, "contabilità della raccolta di informazioni"), che

quantifica il carico annuale di documentazione che il governo impone ai suoi cittadini. In anni recenti, la cifra annuale è vicina ai 10 miliardi di ore! Nonostante cambiamenti significativi, questo carico è elevato da molto tempo. L'ICB presenta anche in dettaglio come è distribuito quel carico fra le varie agenzie e mostra, fra le altre cose, che il Dipartimento del Tesoro è responsabile di ben oltre la metà del totale.

Non dovrebbe essere difficile per i governi di tutto il mondo produrre un ICB, catalogando il peso delle richieste di informazioni ai cittadini. Certo, una parte di quel carico è sicuramente giustificata. Inoltre, le forme peggiori di pantano potrebbero non essere affatto legate alla compilazione della documentazione (pensiamo al tempo che si perde facendo la fila). Per i governi, però, un ICB è un punto di partenza importante, non ultimo perché è probabile che dia il via a sforzi per la riduzione del pantano. Le organizzazioni private dovrebbero produrre documenti simili, se non altro per uso interno; e la trasparenza pubblica sarebbe una buona idea.

Le organizzazioni dedicate alla crescita economica, alla salute pubblica, alla buona istruzione, all'aria e all'acqua pulita, all'uguaglianza sessuale, ai diritti di voto, alla riduzione della povertà, così come le piccole imprese in buona salute e le startup non organizzano manifestazioni al grido di "Ridurre il pantano, subito!". Il tempo però è il bene più prezioso che abbiano gli esseri umani. Dovremmo darci da fare per trovare il modo di darne loro il più possibile.

---

1. Si tratta di un'azienda che fornisce servizi di broadcasting e web radio. [NdR]
2. www.whitehouse.gov/briefing-room/statements-releases/2024/08/12/fact-sheetbiden-harris-administration-launches-new-effort-to-crack-down-on-everydayheadaches-and-hassles-that-waste-americans-time-and-money/.
3. J. Tasoff, R. Letzler, "Everyone believes in redempiont: Nudges and overoptimism in costly task completion", in *Journal of Economic Behavior & Organization*, 107, 2014, p. 107. [Il titolo gioca sul doppio significato del termine inglese *redemption* – "riscatto" (di un coupon) e "redenzione" (in senso morale o religioso). In italiano il doppio senso perde un po' di forza. NdT]
4. S. Mullainathan, E. Shafir, *Scarcity. Perché avere poco significa tanto*, tr. it. il Saggiatore, Milano 2014.
5. Vedi il capitolo 3.

6

IL PROBLEMA BARBIE: MANIPOLATI DALLA PRESSIONE SOCIALE

Affrontiamo ora il problema della manipolazione da un punto di vista molto diverso. In questo capitolo non ci occuperemo di influenze nascoste o occulte, che finora sono state al centro della nostra attenzione. Non ci occuperemo di prestidigitazione o di pantano. Ci concentreremo invece sul modo in cui le dinamiche sociali possono portare le persone a compiere scelte che deplorano. Gli esperti di marketing più abili (fra cui anche i funzionari pubblici) spesso manipolano le persone mediante l'uso strategico o lo sfruttamento di quelle dinamiche. Questo è un tipo diverso di inganno.

Il problema fondamentale è che *le persone possono essere manipolate e indotte all'acquisto di beni che vorrebbero non esistessero*. Acquistano quei beni non perché li apprezzino realmente, ma per i segnali sociali che inviano, o per le perdite sociali in cui incorrono, se non li acquistano.

Online, succede ogni giorno. Possiamo chiamarlo il "problema Barbie".

BENI NEGATIVI

Nonostante il successo del bel film del 2023, e con buona pace di chiunque ami realmente Barbie, sospetto che molti bambini, e anche molti genitori, vorrebbero che non esistesse una cosa come la Barbie. (Se non siete d'accordo con il mio sospetto, non importa; per favore, seguitemi comunque.) Questo è vero anche se le bambine giocano con la Barbie, e anche se i genitori acquistano le Barbie, e anche se le bambine chiedono sinceramente una Barbie. Un problema è che non avere una Barbie può

trasmettere un segnale indesiderato. Un altro è che non avere una Barbie può produrre l'esclusione da una rete sociale.

(Non mi piacciono le Barbie. Proprio no. La mia primogenita ne aveva una, e mi faceva venire i brividi. Va bene, è un problema mio.)

Come vedremo, ci sono ovunque Barbie. I nuovi cellulari possono essere Barbie. Molti prodotti online sono Barbie. Mostreremo che anche le piattaforme di social media possono essere a loro volta Barbie, perlomeno per parecchi utenti. I consumatori in generale, e gli utenti delle piattaforme, vengono manipolati dall'uso strategico delle influenze sociali.

IL PIACERE DI NON ESSERCI

Per vedere la struttura del problema, immaginate che per il prossimo sabato sia stata organizzata una festa. Immaginate anche che molti dei vostri amici vi parteciperanno. Immaginate infine di avere due opzioni: (1) andare alla festa; (2) non andare alla festa. Ipotizzando che la festa si tenga, potreste scegliere (1). In effetti, la vostra preferenza può essere chiara e decisa. Potreste essere inclini a pagare molto pur di partecipare alla festa.

Allo stesso tempo, potreste preferire un'altra opzione: (3) la festa non si tiene. A pensarci bene, potreste desiderare che l'idea della festa non fosse stata nemmeno presa in considerazione. Oppure potreste pensare che venga annullata. L'ordinamento delle vostre preferenze è: (3), (1), (2).

Perché questo è un ordinamento immaginabile delle preferenze? Una risposta generale è la FOMO (*fear of missing out*), la paura di perdersi qualcosa, ma questa espressione è ambigua. Che cosa significa? Di solito la si intende nel senso che, se ci si perde qualcosa, c'è qualche probabilità che ci si perda qualcosa di importante – il che fa pensare che dovreste preferire (1) sia a (2) che a (3). Questa non è la situazione su cui mi concentro qui. Perciò rimane l'interrogativo: perché preferire (3) a (1), ma (1) a (2)? Sembra che si stia parlando del "piacere di perdersi qualcosa".

Una possibilità è che, se non andaste alla festa, dareste un segnale che preferireste non dare. Il segnale può essere che non vi piacciono le feste, che non vi piacciono i vostri amici, o che non vi piace la persona che ha organizzato la festa. Il costo dell'invio di qualsiasi di questi segnali può sembrare molto elevato. Potreste mettere in pericolo o perdere le amicizie, oppure compromettere qualche relazione. Se però la festa viene annullata,

potete evitare un evento che preferireste evitare, senza inviare alcun segnale indesiderato.

Va notato che una caratteristica cruciale del ragionamento è il *significato sociale* del non prendere parte alla festa.[1] Potreste non *avere intenzione* di far arrivare quel segnale (i vostri amici vi piacciono, vi piace la persona che ha organizzato la festa), ma avete poco o nessun controllo sul significato che daranno gli altri a quel segnale.

Questo, quindi, è un possibile motivo per preferire (3) a (1) e (1) a (2): il segnale non voluto che viene recepito quando ci si rifiuta di andare alla festa. (Va notato che molto dipende dall'ambiente informazionale in cui si agisce. Se le persone sanno che volete loro bene, ma le feste non vi piacciono, potreste riuscire a dire che non andrete alla festa senza inviare il segnale indesiderato.)

Esiste un'altra possibilità. Potreste pensare che, se la festa si fa, ci saranno occasioni di costruire relazioni. Potreste non avere molto interesse per quelle occasioni; in caso contrario, non vorreste che la festa venga annullata. Comunque, potreste pensare che, se *altre persone sfruttano quelle occasioni e voi no, vi troverete in svantaggio comparativo*.[2] Questa è una concezione della *fear of missing out*.

Se la festa viene annullata, non dovrete parteciparvi, che è una buona cosa (ottima!) e non sarete in svantaggio, che è un'altra buona cosa (ottima!). Andreste alla festa per evitare quello svantaggio, ma se la festa non si svolge sarà ancora meglio. Notate che chi organizza la festa potrebbe manipolare le persone perché vi partecipino sfruttando i danni sociali che subirebbero gli invitati riluttanti se decidessero di non parteciparvi. (Parleremo di consumo e social media al momento opportuno.)

Ora supponiamo che andiate alla festa e che circolino bevande alcoliche. Avete due possibilità: (1) bere o (2) non bere. Potreste scegliere (1), anche se preferiste se (3) non venissero servite bevande alcoliche alla festa. Potreste scegliere (1) rispetto a (2) per lo stesso motivo per cui avete deciso di andare alla festa. Un possibile problema è il segnale di (2). Potreste temere che le persone si chiedano: siete alcolisti? Siete puritani? Non vi piace divertirvi? Perché non vi lasciate andare?

(Mia moglie è irlandese e ogni anno andiamo in Irlanda. A me non piace bere. Se mi offrono da bere, dico di no e riesco intuire quando i parenti irlandesi pensano: "Magari è un alcolista!". Non è bello. Oppure pensano:

“Forse disapprova l’alcol. Forse è un rompiscatole moralista”. Ed è ancora peggio.)

Supponiamo ora che siate nel bel mezzo di una conversazione. Non è un granché come conversazione. Non è una vera sofferenza; è abbastanza piacevole, anche se vi state annoiando e siete un po’ irrequieti. Avete due possibilità: (1) continuare la conversazione e (2) porre fine alla conversazione e spostarvi. Scegliete (1), anche se preferireste di gran lunga se (3) l’organizzatore della festa vi interrompesse, così da non avere altra scelta se non interrompere la conversazione.

Potreste scegliere (1) rispetto a (2) per lo stesso motivo per cui avete deciso di andare alla festa. In effetti spesso le persone continuano una conversazione più a lungo di quello che desidererebbero, forse soprattutto perché porre fine a una conversazione invia un segnale che le persone non vogliono mandare.[3]

BENEFICENZA, POLITICA E ALTRO

Quello che vale per gli eventi puramente sociali vale anche per le attività di beneficenza o di natura politica. Le persone possono essere manipolate in modo da dedicare tempo o denaro. Può essere una cosa accettabile, anche una buona cosa, ma potrebbe anche non esserlo.

Supponiamo che un martedì sera ci sia un evento per una raccolta fondi o una protesta di qualche tipo. Potreste partecipare all’uno o all’altra, magari perché non volete dare il segnale che la causa pertinente non vi sta a cuore. Nel caso estremo, in realtà proprio quella causa non vi sta a cuore, o magari addirittura la disapprovate.

I costi personali della mancata partecipazione, però, possono sembrarvi molto elevati. Potreste desiderare che l’evento non sia stato organizzato o che venga annullato. Se ha luogo, però, voi siete lì. Potreste non volere segnalare ai vostri amici che quell’ente di beneficenza o quella causa non vi stanno a cuore. Potreste pensare: “Voglio che quell’ente di beneficenza prosperi, ma non voglio partecipare all’evento o dare un contributo in denaro. Se però l’evento ha luogo, è meglio che partecipi e dia un contributo”. (Magari c’è un amico che vi spinge con forza a farlo.)

Oppure, in casi del genere, potreste concentrarvi sull’autosegnalazione; non volete segnalare, soprattutto a voi stessi, che non avete molto a cuore

una certa causa che, pensandoci bene, approvate. Che immagine avrete di voi, se date a voi stessi quel segnale imbarazzante o addirittura vergognoso?

Va notato che, in casi del genere, è possibile che il desiderio delle persone di non inviare un segnale indesiderato ad altri o a se stesse risolvae un problema di azione collettiva: *il gruppo trae beneficio dall'ampia partecipazione, anche se il singolo ha un incentivo a defezionare (e a stare a casa)*. Può essere in atto una norma sociale benefica: bisogna partecipare a eventi che riguardano attività di raccolta fondi o a proteste rilevanti. Pensandoci bene, potreste onorare quella norma, anche se, in qualche senso, desidereresteche l'evento non si verifichi. Torneremo su questo punto.

Supponiamo ora che non siate contenti del vostro lavoro e che abbiate un'altra opportunità che preferireste. Supponiamo che vi piacciano il vostro capo e i colleghi, e che li ammiriate, il che significa che siete molto riluttanti a licenziarvi; vi sentireste in colpa e un po' vi vergognereste. Non vorreste troncare i rapporti con loro, anche se vi piacerebbe poter schiacciare un pulsante magico e trovarvi a lavorare altrove. Fra licenziarvi e continuare a lavorare nello stesso posto, preferite continuare a lavorare lì, per come vi sentireste se vi licenziaste. Ma se la vostra posizione venisse improvvisamente eliminata, per ragioni indipendenti dalle vostre prestazioni – se non aveste l'opzione di mantenere lo stesso posto di lavoro –, sareste ancora più felici.

L'ordinamento delle vostre preferenze potrebbe essere questo: (1) la vostra posizione viene eliminata; (2) rimanete nello stesso posto di lavoro; (3) vi licenziate. (Per ovvie ragioni, non volete l'opzione (4), che è essere licenziati.) Qui siete concentrati sul segnale che inviate licenziandovi, e anche sulla potenziale rottura dei rapporti personali.

Immaginate ora che ci sia una competizione di qualche genere, e che potreste essere i vincitori. Supponiamo che speriate davvero di vincere. Supponiamo che paghereste molto per vincere, e anche che paghereste qualcosa per avere una probabilità di $1/x$ di vincere, anche se x è un numero molto elevato. Anche così, è possibile che preferireste che quella competizione non esistesse. Il motivo è che è probabile che perdiate e, se perdete, sarete molto tristi. Può essere che tutte o quasi tutte le persone che potrebbero ottenere il primo premio si sentano come voi. Il problema di fondo ha a che vedere con la competitività: qualcuno odia perdere, il che significa che neanche gli piace entrare in competizione; e, se c'è una competizione, vuole assolutamente vincere.

Ora supponiamo che qualcuno si sposi, in una località molto lontana, che abbia invitato alla cerimonia molte persone, fra cui anche voi. L'invito non vi fa piacere. Magari non vi va l'idea del viaggio, magari non vi piacciono i matrimoni (in particolare quelli con un gran numero di invitati). Potreste pensare: per me, i costi dell'andare a quel maledetto matrimonio superano di gran lunga i benefici.

Potreste comunque andare. Potreste pensare che, se non andate, manderete un bruttissimo segnale agli sposi e, magari, anche a molti altri. Potreste desiderare di non essere stati invitati a quella cerimonia – o che fosse stato organizzato un matrimonio con pochi invitati, fra cui non voi. Per un esempio online, il messaggio mostrato nella figura 6.1 può essere visto come evidenza della "paura di perdersi qualcosa" degli utenti di Spotify.

So di aver presentato molti esempi. Il mio obiettivo è mostrare la struttura del problema, e la varietà delle Barbie. Ovviamente gli esempi non sono identici, ma hanno ampie zone di sovrapposizione. La caratteristica comune è che riguardano qualche bene che le persone consumano, o qualche attività a cui le persone si dedicano, anche se desidererebbero che quel bene o quell'attività non esistessero. Le ragioni dell'ordinamento delle loro preferenze variano. Il problema può riguardare la segnalazione ad altri; le reti sociali e lo svantaggio comparativo; l'autosegnalazione; la competitività. (Non pretende di essere un elenco esaustivo.) Sullo sfondo, può esserci un manipolatore che tira i fili come un burattinaio – non nascondendo qualcosa, o facendo appello al Sistema 1, ma sfruttando le influenze sociali. Parliamo, in sostanza, di un tipo di trappola.

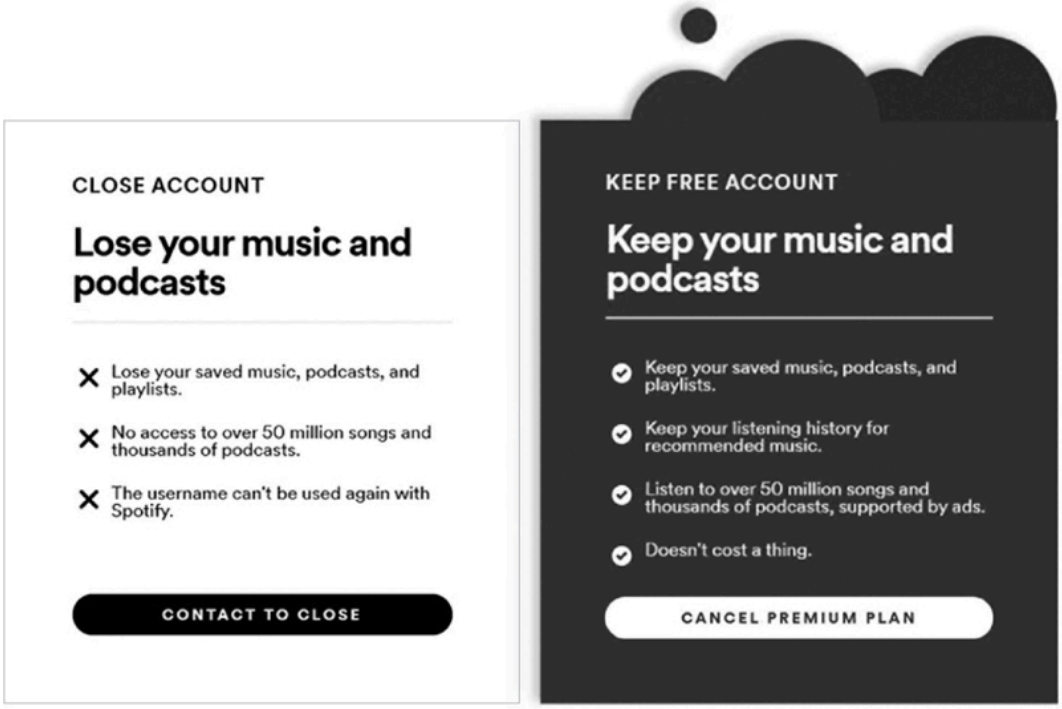

Figura 6.1 Spotify e la paura di perdersi qualcosa.

Fonte: Deceptive Patterns, Hall of Shame, www.deceptive.design/hall-of-shame.

BARBIE, MANIPOLAZIONE E TRAPPOLE DEI PRODOTTI

Molti studi cercano di capire quanto le persone valutino le piattaforme di social media. Se alle persone negli Stati Uniti viene chiesto quanto chiederebbero per rinunciare (poniamo) a Instagram o YouTube per un mese, parlano di una cifra significativa, spesso intorno ai 50 e 100 dollari.[4] Mi sono dedicato anch'io a questi studi, e ho appurato che, per molte piattaforme, la domanda media è proprio intorno a quelle cifre. Risultati di questo tipo fanno pensare che le piattaforme di social media producano un sostanziale surplus per i consumatori. Le persone sono molto fortunate che esistano quelle piattaforme! Ricevono, gratuitamente, un prodotto che per loro vale denaro sonante.

Instagram ha circa 2,4 miliardi di utenti in tutto il mondo. Se poniamo che l'utente medio chiederebbe 10 dollari per non accedere a Instagram per

un anno (una stima molto prudenziale), vorrebbe dire che la piattaforma genera una cifra incredibile di benefici ogni anno, 20,4 miliardi di dollari.

Potremmo però essere un po' cauti, prima di accettare questa conclusione. Le persone potrebbero chiedere molto per non accedere per un mese a una piattaforma di social media, *supponendo che tutte le altre persone che sono rilevanti per loro utilizzino quella stessa piattaforma nel corso di quel mese*. Ma questo che cosa ci dice, esattamente? Può dirci che le persone traggono beneficio dall'essere sulla piattaforma, nel senso che questa fornisce loro vari beni. Può dirci che le persone traggono piacere o comunque qualche beneficio dall'essere parte di una rete. Ma ci dice che sono felici che quella piattaforma esiste? Se potessero premere un pulsante e abolire quella piattaforma, lo farebbero se la loro unica preoccupazione fosse il loro benessere?

Leonardo Bursztyn e i suoi collaboratori hanno cercato di rispondere a questa domanda.[5] Hanno scoperto che, mediamente, gli utenti chiederebbero 59 dollari per disattivare TikTok per un mese, e 47 dollari per disattivare Instagram per lo stesso periodo. Da questi risultati, che sono sostanzialmente coerenti con quelli ottenuti da altri studi, potremmo concludere che le due piattaforme fanno stare molto meglio le persone. In netto contrasto, però, gli utenti sarebbero disposti a *pagare* 28 dollari perché tutto il mondo, loro inclusi, disattivasse TikTok per un mese – e 10 dollari per fare la stessa cosa con Instagram.

Quasi un terzo degli utenti attivi di TikTok sembra perdano benessere per l'esistenza della piattaforma! Lo stesso vale per quasi metà degli utenti attivi di Instagram.

Sembra dunque che *molte persone chiederebbero una cifra significativa per smettere di utilizzare un prodotto che vorrebbero non esistesse*. Cosa degna di nota, c'è dell'eterogeneità in questi conteggi; alcune persone ottengono veramente un vantaggio dall'esistenza di entrambe le piattaforme, ma un numero molto grande di persone crede invece di perdere qualcosa.

È importante dire che una previsione *ex ante* degli effetti che può avere la disattivazione sul benessere può non misurare gli effetti reali.[6] Non sappiamo mai in anticipo se un bene o un'attività ci miglioreranno la vita. Ho sottostimato i benefici di avere un cane. (È una cagnolina, il suo nome è Snowy.) Ho sovrastimato i benefici di avere un ventilatore in più a casa. (Un ventilatore basta.) Le persone possono essere disposte a pagare

qualcosa per un bene a causa del quale in ultima istanza perdono benessere, anche al di là degli effetti di rete. Le persone possono non essere disposte a pagare qualcosa per un bene da cui potrebbero guadagnare qualcosa in termini di benessere, anche indipendentemente dagli effetti di rete.

Se però le persone sono propense a chiedere un pagamento per rinunciare a un bene quando sarebbero disponibili a pagare anche per la sua abolizione, abbiamo un buon motivo per pensare che le persone simultaneamente traggano beneficio dall'avere accesso a quel bene, data la sua esistenza, e trarrebbero beneficio dalla sua eliminazione, se solo fosse possibile. Bursztyn e collaboratori hanno raccolto dati empirici robusti a sostegno di questa conclusione.[7]

IL CORAGGIO DI ULISSE

Se avete paura di poter acquisire una dipendenza da un certo prodotto, potreste desiderare che quel prodotto non esista. Potreste non voler proibire le sigarette, perché siete convinti che le persone abbiano il diritto di fumare. Riflettendoci, però, potreste pensare che sarebbe meglio se le sigarette non esistessero. Potreste pensare che il mondo sarebbe migliore se non esistessero eroina e cocaina. Potreste pensare, più in generale, che le persone possono essere indotte in tentazione, e che sarebbe bene se alcune o molte fonti di tentazione non fossero mai state inventate.

Nel capitolo 2, abbiamo incontrato la situazione di Ulisse e le sirene, in cui una persona compie attivamente dei passi per proteggersi da una tentazione irresistibile. In qualche caso, le persone non vogliono essere esposte a certi allettamenti, perché sanno di non potervi resistere facilmente; vorrebbero che quelle tentazioni non esistessero, e si comportano in modo da ridurre il rischio di caderne vittima. Senza entrare nei dettagli, si tratta di casi rilevanti. Gli allettamenti possono essere forme di manipolazione.

Questo però, decisamente, non è il problema Barbie. Nei casi su cui mi sono concentrato, le persone non dicono che vorrebbero impegnarsi in anticipo a non fare certe scelte. *Vogliono* fare quelle scelte. Il problema non è la loro avventatezza o la loro impulsività, o qualche forma di debolezza della volontà. Non abbiamo a che fare con una disparità fra il "pianificatore" interno delle persone, che cerca di organizzare una vita, e

l'"esecutore" interno, che intraprende le azioni di momento in momento. Il problema è che non avere accesso al bene pertinente crea una perdita, *dato che quel bene è disponibile ad altri*.

I risultati centrali qui sono in un certo senso l'immagine speculare di un'ipotesi plausibile sulla disponibilità a pagare per proteggere l'ambiente.[8] Supponiamo che alle persone venga chiesto quanto sarebbero disposte a pagare per salvare una specie a rischio di estinzione o un'area incontaminata. Supponiamo inoltre che la cifra sia molto piccola e che, anche calcolata su una popolazione molto numerosa, rimanga relativamente piccola. Potremmo pensare che le persone non abbiano particolarmente a cuore il bene in questione. È possibile però che la gente sia disposta a pagare poco per proteggere l'ambiente *a meno che non abbia la sicurezza che anche altri pagheranno allo stesso modo*. Le persone possono pensare: "Quello che pago io farà poco o niente. Se però tutti pagano quella cifra, anche se è poca cosa, possiamo ottenere molto. Sarei disponibile a pagare di più, se tutti gli altri facessero altrettanto".

LA CACCIA AL CERVO

Torniamo all'esempio con cui ho iniziato e supponiamo che qualcuno abbia deciso di organizzare una festa per questo sabato sera. Supponiamo che le persone siano disposte a partecipare, dato che la festa si terrà, ma che tutte o la maggior parte delle persone preferirebbero che la festa non ci fosse. I partecipanti ambivalenti devono affrontare un problema di azione collettiva che va sotto il nome di "caccia al cervo".[9] Se non possono coordinarsi, e non possono trovare un modo per accordarsi e rimanere a casa, andranno alla festa. Se possono coordinarsi e concordare di non andare alla festa, staranno meglio (a meno che il segnale inviato dalla loro azione collettiva abbia dei costi; magari chi organizza la festa ha un certo potere su di loro). Ma come possono coordinarsi? Il coordinamento può essere molto difficile, in particolare se le persone invitate alla festa sono molte. Potrebbero non riuscire a raggiungere degli accordi credibili le une con le altre.

Ecco il punto centrale: le aziende possono mettere i consumatori in situazioni analoghe a quella della festa indesiderata. Questa è una forma di

manipolazione. In effetti è una forma onnipresente, anche se non ci accade spesso di notarla.

Supponiamo, per esempio, che ci sia un prodotto che i consumatori vogliono avere, data la sua esistenza, ma della cui esistenza non sono felici e che potrebbero addirittura deplorare. Se così fosse, quel prodotto sarebbe simile alla festa del sabato sera. Le aziende potrebbero commercializzare un prodotto sottolineando uno fra i due punti seguenti: (1) le persone si vergogneranno o non si sentiranno a posto se non lo hanno (a causa del segnale negativo collegato al fatto di non averlo) o (2) le persone perderanno parte del loro status in conseguenza del non far parte di una rete. Moltissimi prodotti rientrano in questa categoria: automobili, telefoni cellulari, abbigliamento o laptop. Prodotti di questo genere possono rivelarsi delle Barbie.

Prendiamo le cravatte: molti uomini le indossano, anche se non ne ricavano alcun piacere e potrebbero desiderare che non siano mai state inventate. Le cravatte sono Barbie. Oppure prendiamo le scarpe con i tacchi alti: le donne possono indossarle anche se le trovano scomode e desidererebbero che non esistessero. Gran parte degli oggetti di moda possono essere caratterizzati come Barbie. Cravatte e scarpe con i tacchi alti sono oggetti che molte persone acquistano ma al contempo desidererebbero che non esistessero.

In alcune comunità, le norme rispetto a entrambi questi oggetti sono cambiate, e con una rapidità incredibile. In alcune di quelle comunità, gli uomini in genere dovevano portare la cravatta, ma poi è arrivata la pandemia da COVID-19 e le persone hanno iniziato a lavorare da casa. Sembrava un po' stupido mettersi una cravatta per farsi riprendere da una videocamera sul laptop, e gli uomini hanno smesso di indossare cravatte. Quando è finita la pandemia, le persone sono tornate nei luoghi di lavoro – e molti uomini non hanno più indossato la cravatta. Si è stabilito un nuovo equilibrio, migliorato.

Oppure consideriamo un nuovo iPhone. Potreste non voler continuare a usare un iPhone di un po' di anni fa, anche se vi trovate molto bene, se sapete che è disponibile un nuovo modello, anche se sareste più felici se il nuovo modello non esistesse. Bursztyn e collaboratori hanno raccolto dati empirici coerenti con questa conclusione.[10] Hanno appurato che le persone preferirebbero lanci di prodotti meno frequenti, anche se acquistano

regolarmente i nuovi modelli, data la frequenza con cui vengono messi sul mercato.

Il motivo per questa distribuzione di preferenze non è autoevidente. Forse le persone sono soddisfatte del prodotto che possiedono e prevedono (con rammarico) la propria impulsività (il problema di Ulisse e le sirene). Forse le persone saranno imbarazzate a farsi vedere dagli altri con un modello meno recente, perciò vorrebbero che il tempo che intercorre fra un modello e il successivo non fosse così breve. I nuovi iPhone possono essere Barbie.

In alcuni di questi casi, la struttura del problema è: una persona P intraprende qualche azione (per esempio, organizza una festa, serve alcolici, mette in commercio un prodotto) che costringe un'altra persona Q a compiere una scelta (partecipare o no, acquistare o no) che rivela informazioni su Q che Q non vorrebbe divulgare, e che impone a Q di decidere se dare un segnale che preferirebbe non dare.[11] In molti casi, il dilemma di Q non sorge accidentalmente. È un prodotto dell'azione furba, maliziosa o orientata all'interesse personale di P. Se P è un datore di lavoro, può chiedere ai suoi dipendenti di dichiarare le proprie inclinazioni politiche, fra cui, per esempio, se vorrebbero sindacalizzarsi. Se P è un governo, potrebbe dare ai cittadini la scelta se fare un giuramento di fedeltà (o rifiutarsi, gentilmente). Sono forme di manipolazione.

NORME

Ho detto che le norme spesso fanno il lavoro della legge; possono impedire alle persone di defezionare da qualche iniziativa da cui il gruppo trae beneficio.[12] Una norma può incoraggiare le persone a contribuire a qualche attività condivisa, per esempio spostare i mobili dell'ufficio, donare il sangue, effettuare una donazione a un'iniziativa di beneficenza, contribuire in un modo o nell'altro alla difesa nazionale. Se la norma non fosse in vigore, le persone potrebbero defezionare. Quando una norma risolve un problema di azione collettiva, fa in modo che le persone stiano meglio, anche se ciascuna preferirebbe defezionare e lasciare che il problema venga risolto da altri. Pensandoci bene, le persone dovrebbero accogliere con soddisfazione la norma.

Questi casi sono importanti, ma non sono ciò che mi interessa qui. Mi concentro sul problema Barbie, i casi in cui le persone, pensandoci, non accettano volentieri la norma rilevante. Non pensano che risolva un problema di azione collettiva. Pensano di essere state manipolate o intrappolate in una situazione che le mette a disagio o che deplorano. In alcuni casi, abbiamo una buona norma o una norma essenziale, non un problema Barbie. Se le persone vanno ai matrimoni o ai funerali a cui desidererebbero non essere state invitate, a ben riflettere potrebbero rendersi conto che molti o la maggior parte di noi traiamo beneficio da una norma; se partecipate al matrimonio di altri, porterete un beneficio alla coppia che si sposa e ai loro amici e contribuirete alla creazione di una norma di reciprocità: anche loro verranno al vostro matrimonio. Se poi non vi sposerete, o nel caso non li inviterete, l'aver partecipato a quelle nozze contribuirà alla creazione di una sorta di collante sociale.

Qualcosa di simile si può dire dei funerali. A molte persone non fa piacere parteciparvi (certamente a me non fa piacere), ma la presenza può dare conforto alle persone in lutto, e tutti traiamo beneficio da una norma che dice che è bene dare conforto a chi è in lutto. I funerali non sono Barbie, anche se preferireste non esservi stati invitati.

BENI POSIZIONALI

Si è discusso molto della differenza fra beni posizionali e non posizionali.[13] Quando un bene è posizionale, il godimento o l'esperienza di quel bene dipende da quello che godono o sperimentano altre persone. Se è non posizionale, il godimento o l'esperienza che una persona ha di quel bene non dipende da quello che godono o sperimentano le altre persone. Nel caso dei beni posizionali, le persone possono essere manipolate. Chi è interessato a far sì che le persone acquistino o facciano una certa cosa può mettere in evidenza quello che altre persone acquistano o fanno. In molti contesti, i consumatori si trovano su un *tapis roulant posizionale*: le loro scelte non li rendono più felici né li fanno stare meglio, e invece servono in gran parte a mantenerli nella stessa posizione nella gerarchia. Si possono manipolare le persone mettendole su un tapis roulant posizionale.

La salute è plausibilmente un bene non posizionale: le persone vogliono essere in salute, non importa se lo siano o meno le altre persone. Non vorrai

avere un tumore o un disturbo cardiaco, anche se molte persone hanno un tumore o un disturbo cardiaco. I veicoli a motore sono plausibilmente beni posizionali; potresti volere quella che è considerata un'ottima auto, o un'auto di lusso, *solo a causa di quello che è il mix corrente di veicoli*. Le automobili sono, fra le altre cose, segnali. Il secondo motivo riguarda l'immagine di sé: anche se nessun altro potesse sapere quale auto guidi, potresti non volerti pensare come il tipo di persona che guida un'auto non tanto buona. Il terzo motivo è l'invidia: puoi provare invidia per le persone che hanno un'auto migliore della tua.

Il quarto motivo, e il più interessante, è che il mix corrente di veicoli costituisce *il quadro di riferimento* rispetto al quale ciascuno misura la propria auto.[14] Se il mix fosse molto inferiore a quello che è ora, potrebbe andarti bene un'auto molto inferiore; non la percepiresti come molto inferiore. Se il mix invece è molto migliore di quello che era (poniamo) vent'anni fa, la tua esperienza di un'auto di quel periodo non sarebbe molto positiva. Potresti avere l'impressione che quell'auto sia disperatamente primitiva, per molti aspetti. È così anche se la tua esperienza di quell'auto, al tempo in cui è stata messa in commercio, era o sarebbe stata straordinariamente positiva. I beni posizionali danno luogo a *esternalità posizionali*: se qualche persona acquista una nuova auto, impone un costo a quelle che non hanno quell'auto.[15]

I beni posizionali appartengono alla categoria dei beni che le persone consumano ma che desidererebbero non esistessero. Certamente questo è vero se le persone consumano beni di quel genere per evitare di inviare un segnale indesiderato, o per evitare una forma di autosegnalazione spiacevole. Il problema del quadro di riferimento deve essere analizzato in modo diverso. È difficile far finta di non sapere quello che si sa, e se sai di una stupenda auto attuale, non sarà facile che desideri che tutti avessero auto meno che stupende di vent'anni fa. È facile manipolare le persone perché pensino che devono acquistare il migliore fra i prodotti più recenti. Un motivo fondamentale è che le si può convincere che i loro beni o le loro attività impallidiscono di fronte a ciò che è disponibile e stupendo.

MANIPOLAZIONE E DIPENDENZA

Alcune persone hanno problemi di dipendenza. Possono essere dipendenti da sigarette, alcol, eroina o social media. La caratteristica tipica di una dipendenza è che le persone sperimentano (1) minori benefici dal consumo con il passare del tempo e (2) costi più elevati del non consumo con il passare del tempo.[16] La prima volta che si usa una sostanza stupefacente, si può provare una "botta" fantastica e, se non si diventa dipendenti, non si soffrirà se non si usa quella sostanza il giorno dopo. Il problema è che, per alcune sostanze, la "botta" è inferiore la seconda volta (o bisogna assumerne di più per avere la stessa esperienza iniziale) – e che un giorno senza assumerla può diventare disperato o miserabile, perché la si brama fortemente. Il costo in termini di benessere è dovuto al fatto che le persone soffrono sempre di più per il mancato consumo, il che significa che, riflettendoci, desidererebbero non avere iniziato mai a consumare quel bene.[17] Le persone potrebbero addirittura desiderare che quel bene non sia mai *esistito*, anche mentre lo consumano. Come Ulisse, potrebbero adottare una strategia di impegno, per privarsi della possibilità di accedere a beni che liberamente sceglierebbero.

Quanti vendono beni che danno dipendenza possono essere visti come manipolatori. Possono mettere in evidenza la "botta" iniziale, quale che ne sia la fonte. Possono nascondere o passare sotto silenzio le conseguenze della dipendenza. Le piattaforme di social media non sono esattamente sostanze stupefacenti, ma quelli che cercano di fare in modo che le persone ne diventino dipendenti sono molti bravi nel loro lavoro.

Tutto questo è vero e importante, e ha implicazioni forti per la legge. Per esempio, va a sostegno di una proibizione del fumo,[18] magari limitata alle persone nate dopo un certo anno,[19] riconoscendo l'esistenza attuale di dipendenza e per impedire che si diano ulteriori casi. A un certo punto, la Nuova Zelanda ha adottato questa strategia, dicendo che le persone nate dopo una particolare data non avrebbero potuto acquistare sigarette. Anche se la legge non è stata approvata, si potrebbe dire molto in suo favore. Io penso che fosse un'idea eccellente.

Le persone consumano beni che danno dipendenza, anche se vorrebbero che quei beni non fossero mai esistiti. La dipendenza è parente del mio interesse primario qui, ma è solo quello. Non è un problema Barbie.[20] Il

motivo è che, una volta che le persone hanno acquisito una dipendenza, non desiderano che la fonte della loro dipendenza non esista. La desiderano fortemente. Certo, vorrebbero non esserne dipendenti, e potrebbero desiderare che il bene da cui sono dipendenti non fosse mai stato creato, ma questi sono problemi di altro tipo. Esiste un'argomentazione robusta a favore della regolamentazione dei beni che danno dipendenza, di qualsiasi tipo siano, se e perché riducono il benessere delle persone, e le rendono meno libere.

POLITICA E LEGGE

Quali sono le implicazioni per la politica e la legge? Per i beni che danno dipendenza, la risposta è relativamente ben compresa. Le persone possono essere informate o avvisate; alle persone si può dare una spinta gentile; i beni rilevanti possono essere tassati;[21] i beni rilevanti possono essere sottoposti a limitazioni[22] o addirittura messi fuori legge. Anche per i beni posizionali la risposta è ben compresa. Questi beni possono essere tassati,[23] oppure si possono pungolare le persone perché non li acquistino. (Parleremo più a fondo dei pungoli nei capitoli 7 e 8.) Si può pensare che una tassa progressiva sul consumo sia la risposta migliore alle esternalità posizionali.[24] Può ridurre la manipolazione che può essere indotta dalla competizione posizionale.

E il problema Barbie? Che cosa dire dei beni che le persone consumano per non dare un segnale indesiderato, o per non essere svantaggiate in qualche rete o in qualche competizione di status? Dovrebbe essere chiaro che, finché i beni rilevanti esistono, un pungolo, un avvertimento o una tassa possono non essere una risposta adeguata, e potrebbero non essere nemmeno una buona idea. Le persone dovrebbero essere avvertite a proposito di Barbie, cravatte o tacchi alti? Perché avvertire le persone di non andare alle feste a cui (razionalmente) vogliono partecipare, data l'esistenza di quelle feste? Barbie, cravatte o tacchi alti dovrebbero essere tassati?

Una strada per risolvere il problema sta nell'azione volontaria da parte delle persone che sono ostacolate o intrappolate, magari attraverso pressioni dei consumatori o la creazione di una norma di qualche tipo. Forse il problema della caccia al cervo può essere risolto con qualche tipo di

cooperazione.[25] Le persone possono essere in grado di stringere un accordo e rispettarlo. So di qualcuno che nel governo degli Stati Uniti ha adottato una norma: tutte le riunioni saranno limitate a quindici minuti. Molte persone l'hanno apprezzata. Mia sorella ha proposto una regola per la nostra famiglia a Natale: niente regali per gli adulti, mai. Tutta la famiglia apprezza quella norma.

Di fronte a certi tipi di manipolazioni, che producono il problema Barbie, potremmo immaginare casi in cui abbia senso considerare un'azione legale. Ricordiamo: c'è la possibilità che qualche agente abbia deliberatamente predisposto una situazione in cui le persone devono o (a) consumare un prodotto che non amano o intraprendere un'attività che non fa loro piacere o (b) offrire un segnale che preferirebbero di gran lunga non dare, o divulgare un fatto che vorrebbero mantenere privato. Se il risultato è una seria perdita di benessere o l'intrusione in diritti protetti, l'intervento legale, magari sotto forma di pungoli, deve essere preso in considerazione.

Un esempio può essere un caso in cui un ente pubblico dà ai suoi dipendenti la possibilità di sostenere un politico, o di partecipare a una celebrazione festiva associata a una specifica tradizione religiosa, o di pregare insieme il lunedì mattina. Se alcune azioni di questo tipo producono un danno, l'intervento legale non deve essere escluso – magari sotto forma di pungoli, magari sotto forma di tasse, magari sotto forma di sussidi e magari sotto forma di obblighi e messe al bando. Negli Stati Uniti, si può vedere lo Hatch Act come una risposta specifica a un problema Barbie: proibisce a certi dipendenti pubblici di impegnarsi in attività politiche, in parte in base all'idea che, se venisse chiesto loro di farlo da parte di funzionari di alto livello, quelle persone non potrebbero rifiutarsi facilmente. Sarebbero in trappola.

FUORI DALLE TRAPPOLE

Il problema è di grandi proporzioni e onnipresente. Le Barbie sono comuni. Chi agisce nel proprio interesse, online o meno, manipola le persone sfruttando le influenze sociali. In campi importanti, sembrerebbe ci sia una base per ripensare la legge sulla protezione dei lavoratori e dei consumatori e per prendere in considerazione una potenziale azione. In effetti possiamo andare oltre. Nel caso dei social media, che qui ho messo

in evidenza, molte persone affrontano un problema Barbie, il che va a favore di riforme politiche, magari soprattutto per aiutare le persone più giovani a evitare le trappole dei prodotti. Interventi privati e pubblici, pensati per aiutare le persone a sfuggire a quelle trappole, possono aumentare il benessere sociale; pensiamo agli sforzi per pungolare gli adolescenti a non trascorrere del tempo sulle piattaforme di social media a tarda notte. Se poi alle persone viene chiesto di rinunciare ai loro diritti, sarebbe meglio proibire la richiesta di rinuncia. Le richieste di rinuncia possono essere Barbie.

La forma di manipolazione che ho esaminato qui è molto generale. Se le persone possono coordinarsi o stringere qualche tipo di accordo, potrebbero essere in grado di trovare un modo per proteggersi dalle azioni dannose, compreso l'acquisto di prodotti che desidererebbero non esistessero. Coordinamento e accordo però potrebbero non essere possibili. I venditori di vario tipo contano su questo. È una forma ben precisa di manipolazione, e bisognerebbe rimediarvi.

---

1. Vedi L. Lessig, "The regulation of social meaning", in *University of Chicago Law Review*, 62, 1995, p. 943.
2. Cfr. L. Bursztyn et al., "Status goods: Experimental evidence from platinum credit cards", in *Quarterly Journal of Economics*, 133, 2018, p. 1561; R. Mason, "Measuring the demand for status goods: An evaluation of means-end chaining and laddering", in *European Advances in Consumer Research*, 2, 1995, p. 78.
3. A.M. Mastroianni et al., "Do conversations end when people want them to?", in *PNAS*, 118, 2021, p. 1.
4. Vedi H. Allcott et al., "The welfare effects of social media", in *American Economic Review*, 100, 2020, p. 629; C.R. Sunstein, "Valuing Facebook", in *Behavioural Public Policy*, 4, 2020, p. 370. Va notato che questi studi chiedono quanto chiederebbero le persone per non utilizzare le piattaforme, non quanto pagherebbero per poterle utilizzare. Nella mia ricerca venivano poste entrambe le domande e il risultato era che la cifra che le persone erano disposte a pagare per l'uso era molto inferiore; ho esplorato i motivi della differenza nello stesso studio.
5. L. Bursztyn et al., "When product markets become collective traps: The case of social media", NBER Working Paper 31771, 2023.
6. D.T. Gilbert, T.D. Wilson, "Miswanting: Some problems in the forecasting of future affective states", in J.P. Forgas (a cura di), *Feeling and Thinking: The Role of Affect in Social Cognition*, Cambridge University Press, Cambridge (UK) 2001, p. 178; D.T. Gilbert et al., "Immune neglect: A source of durability bias in affective forecasting", in *Journal of Personality & Social Psychology*, 75, 1998, p. 617.
7. L. Bursztyn et al., *op. cit.*
8. Vedi A. Sen, "Environmental evaluation and social choice: Contingent valuation and the market analogy", in *Japanese Economic Review*, 49, 1995, p. 23.

9. Vedi B. Skryms, *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, Cambridge (UK) 2004.

10. Vedi L. Bursztyn et al., *op. cit.*

11. Ringrazio Eric Posner per avermi fatto riflettere su questo aspetto.

12. Vedi E. Ullmann-Margalit, *The Emergence of Norms*, Clarendon Press, Oxford 1977.

13. Vedi R.H. Frank, "Positional externalities cause large and preventable welfare losses", in *American Economic Review*, 95, 2005, p. 137; R.H. Frank, "The demand for unobservable and other nonpositional goods", in *American Economic Review*, 75, 1985, p. 101.

14. Vedi R.H. Frank, *Luxury Fever*, Princeton University Press, Princeton 1999.

15. Vedi R.H. Frank, "Positional externalities…", cit.

16. C.R. Sunstein, "Legal interference with private preferences", in *University of Chicago Law Review*, 53, 1986, pp. 1129, 1159.

17. Si può discutere di questa formulazione. Dobbiamo sapere di più sui benefici e i costi del consumo, al passare del tempo, per sapere se le persone, riflettendoci, desiderino di non avere mai iniziato. Pensate a una storia romantica con un grande inizio, un periodo di dipendenza, e una conclusione terribile.

18. Vedi R.E. Goodin, *No Smoking*, University of Chicago Press, Chicago 1989.

19. Per un esempio di questa politica, vedi J. Gretener, K. Magramo, "New Zealand bans tobacco sales for next generation", CNN, 13 dicembre 2022, https://www.cnn.com/2022/12/13/asia/new-zealand-tobacco-ban-intl-hnk/index.html.

20. Intendo dire che il consumo di beni che danno dipendenza non è motivato da (1) un desiderio di evitare l'invio di segnali o (2) un desiderio di evitare svantaggi comparativi dovuti al mancato consumo di beni che danno dipendenza, mentre tutte le altre persone li consumano. Perlomeno, speriamo che non sia così.

21. Vedi A.I. Friedson et al., "Cigarette taxes, smoking, and health in the long-run", NBER Working Paper 29145, 2021.

22. Vedi D.M. Cutler et al., "Economic approaches to estimating benefits of regulations affecting addictive goods", in *American Journal of Preventive Medicine*, 50, 2016, p. S20.

23. Vedi A. Antinyan et al., "Curbing the consumption of positional goods: Behavioral intervention versus taxation", in *Journal of Economic Behavior & Organization*, 179, 2020, p. 1; vedi anche R.H. Frank, "Positional externalities…", cit. (dove questa idea è discussa ma respinta per motivi di amministrabilità).

24. Vedi R.H. Frank, "Positional externalities…", cit.

25. Un suggerimento di azione collettiva, dai consumatori, è una lettura corretta di L. Bursztyn et al., *op. cit.*, alla luce delle significative perdite di benessere.

PARTE TERZA

IL FUTURO

7

PUNGOLI
SENZA MANIPOLAZIONE, 1

Ogni giorno, le persone vengono influenzate senza essere manipolate. Potreste incontrare un avvertimento: “Questo prodotto contiene crostacei”, oppure “Il fumo causa il cancro”. Potreste ottenere informazioni: “Gli adulti hanno bisogno di 2000 calore al giorno”. Potreste ricevere un promemoria: “Appuntamento presso il reparto di medicina interna il 10 giugno alle 10.00”. Avvertimenti, informazioni e promemoria non sono per forza manipolativi. Oppure potreste essere automaticamente iscritti a qualcosa – l’energia verde, per esempio. È un caso di manipolazione? Come abbiamo visto, la risposta non è ovvia.

Molte persone, fra cui chi scrive, si sono interessate ai “pungoli”, che definiremo subito più precisamente, e alla “architettura delle scelte” come strumenti di regolamentazione.[1] Nel capitolo 2, abbiamo visto che alcuni pungoli superano il confine e vanno considerati forme di manipolazione. L’obiettivo principale di questo capitolo, e di quello seguente, è analizzare se e quando i pungoli sollevano problemi etici gravi. Per fare qualche passo avanti a questo proposito, continueremo a esplorare la manipolazione, ma amplieremo un po’ lo sguardo e prenderemo in considerazione gli aspetti etici più in generale.

Come vedremo, i problemi ruotano in gran parte intorno al fatto che i pungoli promuovano o invece minino il benessere, l’autonomia e la dignità. Come vedremo, l’analisi etica dei pungoli è simile all’analisi corrispondente per altri strumenti, come le multe e gli obblighi, ma esistono sfumature specifiche, che riguardano potenziali minacce all’autonomia e

alla dignità. In particolare, le obiezioni più plausibili ai pungoli sono motivate da una preoccupazione per la capacità di agire personale, che perciò qui sarà un tema unificante.

La mia conclusione centrale è che, almeno se vengono prese in generale o in astratto, le obiezioni etiche non hanno molta forza, per due ragioni diverse. La prima: sia pungoli sia architettura delle scelte sono inevitabili, e perciò non ha alcun senso desiderare che scompaiano. La seconda: molti pungoli e molte forme di architettura delle scelte mantengono e promuovono la capacità di agire. Non manipolano nessuno. Per esempio, la divulgazione di informazioni ("questo prodotto contiene crostacei") non è per forza manipolativa. Molti pungoli sono difendibili e addirittura potrebbero essere indispensabili per motivi etici, non importa se ciò che ci sta a cuore sono benessere, autonomia, dignità o qualche altro valore.

È vero che ogni azione del governo, compresi i pungoli, deve sostenere l'onere della giustificazione (che a volte può essere un onere pesante). Se il governo richiede la divulgazione di informazioni o iscrive automaticamente le persone in un piano di assistenza sanitaria, deve spiegarsi e difendersi. Il fatto che le persone conservino la libertà di scelta, e che in ultima istanza abbiano la possibilità di andare per la propria strada, è importante, ma non dà ai funzionari pubblici la licenza di fare qualsiasi cosa vogliano. I manipolatori non eliminano la libertà di scelta. In molti casi, però, la spiegazione richiesta è disponibile.

Supponiamo, per esempio, di essere convinti che l'obiettivo dell'ordinamento sociale (comprese quelle forme per cui è responsabile il governo) sia promuovere il benessere sociale. In tal caso favoriremo i pungoli che promuovono il benessere. Prendiamo, per esempio, un obbligo di divulgazione per le società che gestiscono carte di credito, pensato per promuovere scelte informate, che è probabile raggiunga quell'obiettivo. Non vi è niente di manipolativo o di altrimenti eccepibile in questo.

Oppure supponiamo di credere nell'autonomia e nella dignità individuale. In tal caso favoriremo spinte gentili e architetture delle scelte che promuovono quei valori. Pensiamo, per esempio, a un tentativo di sollecitare le persone a fare la propria scelta a proposito del tipo di piano pensionistico che desiderano, ponendo loro precisamente quella domanda quando entrano nel mondo del lavoro. Non è manipolazione, anche se ovviamente la domanda potrebbe essere espressa in una forma

manipolativa. (“Sì! Voglio pensare a proteggermi per quando sarò più vecchio” o “No, non ho alcun interesse per il mio futuro”.)

Se per noi l’autogoverno democratico è un valore, saremo inclini a sostenere pungoli e architetture delle scelte che non manipolano le persone e aiutano a promuovere obiettivi democratici. Qualsiasi democrazia ha una forma di architettura delle scelte che contribuisce a definire e costituire le sue aspirazioni all’autogoverno. In effetti, una Costituzione può essere vista come una sorta di architettura delle scelte per gli architetti delle scelte. Una società che si autogoverna può benissimo pungolare i suoi cittadini a partecipare al processo politico e a votare. I partiti politici adottano regolarmente simili pungoli, e non è affatto illegittimo che i funzionari pubblici incoraggino le persone a votare. Negli Stati Uniti, per esempio, alcuni stati hanno adottato un sistema di registrazione automatica degli elettori, una forma di architettura delle scelte che, senza alcuna ambiguità, è pensata per promuovere la partecipazione al processo politico attraverso l’esercizio del voto.

Ovviamente, non si devono approvare i pungoli o le architetture delle scelte in astratto o in quanto tali. Rimane possibile che la sfiducia nel governo e la fiducia nei mercati ci portino a minimizzare i pungoli, in base a ragioni welfariste o di autonomia. Alcuni pungoli e alcune forme di architettura delle scelte incappano effettivamente in obiezioni etiche convincenti, soprattutto perché minano il benessere, l’autonomia o la dignità. Possono essere manipolativi. Come abbiamo visto nel capitolo 2, le informazioni sul rischio relativo sono una forma di manipolazione, e gli esiti possono essere presentati in modo da manipolare le persone.

I pungoli possono anche compromettere, anziché promuovere, l’autogoverno. Supponiamo per esempio che una nazione stabilisca una regola di default, secondo la quale, se non indicato esplicitamente altro, si presume che in caso di un’elezione gli elettori siano a favore del leader in carica. Oppure supponiamo che una nazione stabilisca una regola di default, tale per cui, a meno che i cittadini non indichino altrimenti, alla loro morte la loro eredità vada al partito politico più potente. Sono molti i motivi per opporsi a una regola di questo genere, anche se i cittadini hanno la possibilità di dire di no.

Alcuni pungoli hanno *finalità illecite*, e sono spiacevoli per quel motivo. Come vedremo, trasparenza e responsabilità sono salvaguardie indispensabili, e sia pungoli, sia architetture delle scelte dovrebbero essere

trasparenti. Anche in tal caso, però, esiste un rischio di manipolazione, e quel rischio deve essere evitato.

Parleremo di molte cose qui, ma ci sono tre temi unificanti che vanno sempre tenuti presenti. Il primo riguarda il rapporto fra pungoli e capacità umana di agire. Quando sono in atto pungoli, la capacità umana di agire si mantiene (perché non viene eliminata la libertà di scelta) e la capacità di agire si esplica sempre nel contesto di qualche architettura delle scelte. Questa non è una giustificazione sufficiente per i pungoli, ma è importante.

Il secondo tema unificante riguarda la necessità di avere un senso abbastanza ampio della categoria dei pungoli, e un pieno apprezzamento delle differenze fra di essi. Alcuni sono *educativi*, ma altri no; sono *architetturali*. Alcuni pungoli sfruttano o combattono bias comportamentali, altri no, ma anche fra i primi esistono differenze significative. Molti non sono nemmeno verosimilmente manipolativi; alcuni possono superare il limite.

Il terzo tema unificante è la necessità di calare le preoccupazioni etiche negli esempi particolari. Un'osservazione legittima a proposito delle regole sulle impostazioni predefinite può non valere per gli avvertimenti o i promemoria. Un'obiezione etica all'uso delle norme sociali può non essere applicabile alla divulgazione di informazioni. Qui come altrove, l'astrazione può essere una trappola, e questo sarà uno dei miei temi principali.

CONCETTI E DEFINIZIONI

I pungoli (*nudges*) sono interventi che orientano le persone in direzioni particolari ma consentono loro di andare comunque per la loro strada.[2] Poiché preservano la capacità di agire individuale, tutelano la libertà. Un promemoria è un pungolo; lo stesso vale per un avvertimento. Un dispositivo GPS offre pungoli; una spia luminosa che si accende in auto è un pungolo; lo stesso vale per un avviso acustico, che vi dice che la vostra auto si sta avvicinando troppo a un'altra. Per essere considerato un pungolo, un intervento non deve imporre incentivi (o disincentivi) materiali significativi. Un sussidio non è un pungolo; una tassa non è un pungolo; una multa o una condanna al carcere non sono pungoli.

Per essere considerato tale, un pungolo deve lasciare la piena libertà di scelta. Se un intervento impone costi materiali significativi a chi deve scegliere, potrà essere giustificato, ma non è un pungolo. Alcuni pungoli funzionano perché informano le persone; sono educativi. Altri funzionano perché rendono più facili determinate scelte; per esempio, possono ridurre il pantano. Altri ancora funzionano grazie alla forza dell'inerzia e della procrastinazione.

Quando prendono decisioni, le persone lo fanno su uno sfondo costituito da un'architettura delle scelte. Siete stati in un aeroporto recentemente? Forse avrete notato che, per arrivare al gate, gli aeroporti possono essere progettati in modo da dover passare davanti a vari negozi, o addirittura da dover attraversare qualche negozio, passo dopo passo. Vado spesso in Irlanda, e mi piace molto l'aeroporto di Shannon, che è comodo, veloce e semplice. Grande! Ma se andate da Shannon agli Stati Uniti, dovrete passare attraverso un grande negozio, fornito di ogni sorta di prodotti che inducono in tentazione. Recentemente ci sono passato con i miei figli ed è stato estremamente difficile arrivare fino al gate senza fare qualche acquisto. I miei figli erano pungolati. Quella è una architettura delle scelte astuta e, anche se non comporta trucchi o inganni, può essere caratterizzata, verosimilmente, come una forma di manipolazione (vedi la definizione (8) e anche l'analisi di Barnhill, di cui abbiamo parlato nel capitolo 1).

Ogni mensa ha una sua organizzazione e disposizione dei piatti, che influenzeranno ciò che le persone scelgono. Può disporre i piatti sani in una posizione evidente, oppure no. I supermercati hanno una loro architettura, che può essere pensata in modo da promuovere o scoraggiare certe scelte da parte dei visitatori (per esempio, quella di andarsene senza fare un acquisto). Anche se la disposizione di un supermercato fosse puramente casuale o non riflettesse il minimo sforzo di orientare le persone, con tutta probabilità avrebbe delle conseguenze su ciò che la gente finisce per scegliere. Se, per esempio, le persone vedono per primi certi prodotti, è più probabile che li acquistino.[3] Abbiamo chiarito che certe forme di progettazione dei punti vendita possono essere verosimilmente considerate manipolative.

Lo stesso vale per i siti web. Quello che penso sia il miglior libro sul design dei siti web è stato scritto da Steve Krug, e si intitola *Don't Make Me Think*.[4] Il libro è tutto dedicato all'architettura delle scelte e ai pungoli, anche se non utilizza questi termini. La prima legge dell'usabilità di Krug è

già incorporata nel titolo: una pagina web, dice, “deve sembrarmi evidente. Ovvia. Spiegarsi da sé”. Offre tutta una serie di principi pensati per ridurre al minimo lo sforzo cognitivo da parte degli utenti. Krug è un nemico del pantano. Non vuole che gli utenti debbano porre domande. Non c’è nulla di manipolativo in un sito web semplice da usare. Un sito web complicato, invece, può essere stato progettato così proprio per manipolare le persone.

Le istituzioni, private e pubbliche (compresi i tribunali), creano regole di default. In effetti, non possono farne a meno. Un telefono cellulare, un mutuo, un tablet e un programma di welfare inevitabilmente presenteranno delle impostazioni di default per le opzioni, che possono essere modificate se la persona o le persone in questione lo desiderano. Il diritto contrattuale è pieno di regole che stabiliscono che cosa succede se le persone non dicono nulla a proposito di certe opzioni. I rapporti legali delle persone con i loro datori di lavoro, gli istituti che erogano mutui, le società di noleggio auto, le società che erogano carte di credito, e perfino quelli con il partner e i figli sono costituiti in gran parte da regole di default. Pensiamo, per esempio, alle garanzie implicite. Se le parti che contrattano non precisano esplicitamente se una collaborazione possa essere interrotta “arbitrariamente” o “per giusta causa”, la legge deve stabilire un’opzione predefinita; nessuna delle due possibilità è stabilita dalla natura o discende dal cielo. Le regole che stabiliscono valori di default per le opzioni sono pungoli. Spesso si comportano come un dispositivo GPS, o addirittura contribuiscono a dare forma a preferenze e valori.

Per qualcuno è un fatto inquietante o addirittura minaccioso: basandosi sull’idea della capacità di agire individuale, è irritato dai default. Queste persone non sono affatto entusiaste all’idea che aspetti significativi della loro vita siano organizzati da regole di tal genere, che non hanno selezionato in prima persona e che potrebbero derivare benissimo da pratiche, giudizi o desideri altrui. Ciononostante, quell’organizzazione esiste. Inoltre, è vero che alcune regole di default sono un prodotto di tradizioni, usanze, ordini spontanei e mani invisibili – un conforto per chi non ha alcuna fiducia nei funzionari pubblici. Sarebbe stravagante però dire che lo sono tutte. E, anche se lo fossero, pungoleranno le persone che vivono in quel contesto, e ci vuole un bell’impegno per trasformarle in legge, dove avranno effetti significativi.

Anche se una regola di default viene scelta sulla base dell’idea che l’opzione predefinita sia ciò che sceglierebbe la maggior parte delle

persone, e in quel senso è “imitatrice del mercato”, può benissimo avere qualche effetto sulle preferenze e sugli esiti. Se per opzione predefinita la stampante stampa solo sul fronte, i suoi utenti potrebbero finire per preferire la stampa su un solo lato. Una regola di default stabilisce i diritti iniziali delle persone (per esempio, il diritto a essere licenziati solo per giusta causa) e può essere importante per quel motivo, influenzando le preferenze delle persone. Per i nostri scopi qui, il punto è che le regole di default, di un tipo o di un altro, a volte sono inevitabili, o praticamente inevitabili.

Abbiamo visto anche che l’attenzione è una risorsa scarsa. Quando i moduli di richiesta (per i mutui, per le borse di studio, per il rifinanziamento di mutui, per la formazione, per benefici finanziari di qualsiasi tipo) sono complessi, difficili e pieni di pantano, è possibile che le persone non inoltrino la richiesta; come risultato, può andare perduta una grande quantità di denaro.[5] Questo ha implicazioni per la progettazione delle regole. Fa pensare che il settore privato possa aiutare o danneggiare le persone concentrando la loro attenzione in certi modi. Lo stesso vale per il settore pubblico, che tenti o meno di fare altrettanto. Una regolamentazione può essere scritta o applicata in un modo che rende particolarmente salienti certe caratteristiche di una situazione.

Gli “ordini spontanei” sono esaltati da molti.[6] Sono centrali per la tradizione liberale: se un ordine è spontaneo, nel senso che rispecchia le decisioni volontarie, dal basso, di quelli che l’hanno creato, vi è qualche motivo per pensare che rispecchi i giudizi di molte persone su come abbia più senso procedere. In base a certi presupposti, gli ordini spontanei possono promuovere il benessere delle persone, ma sono una forma di architettura delle scelte non meno delle strutture intenzionali, e includeranno un certo numero di pungoli, non ultimo se creano e perpetuano norme sociali (pensiamo alle norme a favore dell’autocontrollo e della sollecitudine). Le mani invisibili possono pungolare tanto quanto le più visibili. Certo, l’ordine spontaneo e le mani invisibili possono essere meno pericolosi delle progettazioni intenzionali, e in base a certi (controversi) presupposti[7] è probabile che siano benigni (o migliori); tuttavia sono comunque forme di architettura delle scelte. Sono manipolativi? Di solito no.

Per il futuro, potremmo immaginare nuove forme di architettura delle scelte pensate per migliorare i programmi di alleviamento della povertà; i

programmi ambientali; i programmi per l’energia; i programmi per il pensionamento e la sicurezza sociale; i programmi antiobesità; i programmi per l’istruzione; i programmi per la sanità; i programmi per aumentare la donazione di organi. Potremmo immaginare anche forme di architettura delle scelte pensate per combattere la discriminazione razziale e sessuale, per aiutare le persone con disabilità e per promuovere la crescita economica. Sarà necessario in futuro dedicare molti studi all’architettura delle scelte in questi e in altri campi.

Comunque, non c’è dubbio che certi pungoli e certi tipi di architetture delle scelte possano sollevare problemi etici gravi. Pensiamo, per esempio, a un governo che usi pungoli per promuovere la discriminazione sulla base di razza, sesso o religione. Qualsiasi governo fascista potrebbe benissimo utilizzare (e quasi certamente lo fa) i pungoli. I terroristi usano pungoli. Persino informazioni veritiere (per esempio, sui tassi di criminalità) possono alimentare le fiamme della violenza e del pregiudizio. (Se le persone vengono a sapere che i cattivi comportamenti sono diffusi, è magari più probabile che adottino quegli stessi comportamenti, perché sono la norma sociale.) Gruppi o nazioni dediti alla violenza a volte adottano pungoli al servizio della loro causa, spesso sotto forma di propaganda, che in certe forme può essere considerata un tipo (negativo) di pungolo. E anche se i pungoli non hanno finalità illecite, è possibile chiedersi se quanti li utilizzano trattano le persone con rispetto.

I maggiori timori per i pungoli e l’architettura delle scelte riguardano tre aspetti fondamentali: benessere, autonomia e dignità. Alcuni pungoli possono andare contro uno o più di questi aspetti, magari perché sono manipolativi. È facile identificare i pungoli che riducono il benessere e conducono le persone a sprecare tempo o denaro; una regola di default poco utile per l’assicurazione sanitaria potrebbe ricadere in quella categoria, e lo stesso vale per una campagna educativa pensata per persuadere le persone ad accendere polizze assicurative eccessive o a fare investimenti stupidi. I pungoli possono essere, e spesso sono, dannosi per l’ambiente. L’inquinamento eccessivo, in parte, è il prodotto di una architettura delle scelte che non aiuta, perché un ampio ventaglio di regole di default e di autorizzazioni legali contribuisce a rendere possibile quell’inquinamento.

L’INEVITABILITÀ DELL’ARCHITETTURA DELLE SCELTE

C’è una bella storiella ripresa dallo scrittore David Foster Wallace: “Ci sono due giovani pesci che nuotano e a un certo punto incontrano un pesce anziano che va nella direzione opposta, fa un cenno di saluto e dice: ‘Salve, ragazzi. Com’è l’acqua?’ I due pesci giovani nuotano un altro po’, poi uno guarda l’altro e fa: ‘Che cavolo è l’acqua?’”.[8]

La storiella riguarda l’architettura delle scelte, che è l’equivalente dell’acqua. L’andamento delle condizioni meteorologiche è a sua volta una forma di architettura delle scelte, perché influisce su ciò che le persone decidono.[9] Non c’è neanche bisogno di dirlo, la vita umana non è immaginabile senza qualche tipo di condizione meteorologica. La natura pungola. La *common law* è un sistema regolatorio, e pungola, anche se lascia alle persone una grande dose di flessibilità.

Da questo punto di vista, l’architettura delle scelte è inevitabile. Gli esseri umani (o i cani, i gatti, i cavalli) non possono sperare di farne a meno. Qualsiasi punto vendita ha una sua organizzazione: certi prodotti si vedono per primi, altri no. Qualsiasi menu colloca le opzioni in varie posizioni. Le emittenti televisive hanno canali con numeri diversi e, cosa che colpisce, i numeri possono essere importanti, anche quando cambiare canale ha un costo estremamente limitato; le persone tendono a scegliere il canale con il numero più basso, cosicché il Canale 3 avrà più spettatori del Canale 53.[10] Ogni sito web ha un suo design, che influirà sulle persone, se, ed eventualmente che cosa, sceglieranno.[11]

Non può evitare i pungoli neanche lo Stato. Supponiamo che un governo sia o sostenga di essere fermamente a favore del libero mercato, della proprietà privata e del *laissez-faire*. Anche in tal caso, non può semplicemente astenersi dall’agire – o dal pungolare. Crea la propria architettura delle scelte. Come scriveva Hayek, il compito di stabilire un sistema competitivo richiede “un campo d’azione ampio e indiscusso all’intervento dello Stato”, poiché “[i]n nessun sistema che possa venir difeso con argomenti razionali si può sostenere che lo Stato non faccia nulla. Un sistema efficace basato sulla competizione necessita, come qualsiasi altro sistema, di un quadro di leggi progettate con intelligenza e continuamente rivisto”.[12]

Come Hayek sapeva, uno Stato che protegga la proprietà privata deve stabilire un insieme di proibizioni e permessi, compreso un insieme di diritti di default, che stabilisce chi possiede che cosa prima che inizi la contrattazione. Ricordate che il diritto contrattuale (come quello relativo alla proprietà e agli illeciti civili) costituisce una forma di architettura delle scelte per l'ordine sociale. Ho notato che i diritti iniziali e le regole di default pungolano perché influiscono sulle preferenze delle persone, anche se i costi di transazione sono bassi e le persone possono contrattare come ritengono meglio. Resta vero che l'architettura delle scelte può preservare la libertà di scelta; è anche vero che gli architetti delle scelte possono almeno aspirare alla neutralità rispetto a dimensioni importanti. L'architettura delle scelte però è inevitabile, il che significa che non ha senso opporvisi sulla base di motivi etici.

LIBERTÀ

Allo stesso tempo, possiamo immaginare questa concezione: alcune forme di architettura delle scelte e di pungoli sono inevitabili, certo, ma è importante ed è una buona cosa se sono il prodotto della natura o di qualche tipo di ordine spontaneo, anziché di un progetto conscio o dell'azione di qualsiasi progettista. Forse la legge può svilupparsi da quell'ordine; forse il diritto civile relativo a contratti, proprietà e responsabilità civile fa esattamente questo. I meccanismi della mano invisibile[13] spesso producono un'architettura delle scelte. Secondo una concezione tradizionale, gran parte del diritto è di fatto "diritto consuetudinario": codifica le pratiche correnti delle persone e non rispecchia alcun tipo di imposizione da parte delle autorità pubbliche. Consideriamo la celebrazione, da parte di Hayek, della "tradizione dell'evoluzionismo empiristico", per la quale

> il valore della libertà consiste principalmente nella possibilità che essa dà allo sviluppo del non intenzionale; e il benefico funzionamento di una società libera è per lo più basato sull'esistenza di istituzioni sviluppatesi spontaneamente. È probabile che un'autentica fede nella libertà non sia mai esistita, e certamente non c'è mai stato alcun tentativo riuscito di creare una società libera, senza un sincero rispetto per le istituzioni, affermata, per le abitudini, gli usi…[14]

In questa concezione, esiste un motivo speciale, dal punto di vista della libertà, per apprezzare forme di architettura delle scelte che rispecchiano il funzionamento di "istituzioni sviluppatesi autonomamente", anziché di

quelle progettate. Possiamo sentirci a nostro agio con qualsiasi pungolo che rispecchi "usi e abitudini", ma sospettare di tutti quelli che non li rispettano. Qui, poi, c'è un fondamento per lo scetticismo a proposito di qualsiasi tipo di ingegneria sociale; lo scetticismo può essere applicato ai pungoli così come a obblighi e proibizioni. Possiamo essere particolarmente preoccupati per l'architettura delle scelte progettata dal governo, in base al fatto che i pubblici funzionari non hanno le informazioni o gli incentivi per essere degni di fiducia.

Anche se il diritto civile relativo a contratti, proprietà e responsabilità civile costituisce forme di architettura delle scelte, e anche se non è molto consuetudinario (e comporta un certo grado di imposizione e di progettazione), l'architettura pertinente può essere resa il più flessibile possibile e mantenere molto spazio per l'ordinamento privato – e quindi per la libertà. Per riassumere una lunga argomentazione:[15] lo Stato, e il diritto, possono fornire le regole di fondo per l'interazione privata e rifiutarsi di specificare gli esiti. Anche se quelle regole si rivelano dei pungoli (come nel caso delle regole di default), sono molto diverse dalla pianificazione sociale in tutte le sue forme – molto più modeste e meno pericolose.

Certo, il diritto penale comprenderà alcune proibizioni; nessuna nazione può affrontare omicidi, rapine a mano armata e stupri con semplici pungoli. Ma forse il diritto penale può limitarsi a proibizioni riguardo all'uso della forza e alle frodi (e più in generale avere un ruolo nel correggere i comuni fallimenti del mercato). Una nazione può sicuramente ridurre al minimo il numero delle attività che criminalizza. Può adottare una concezione ristretta del campo d'azione del diritto penale, magari concentrandosi sui danni ad altri. E può limitare i pungoli a provvedimenti che agiscono al servizio del diritto civile, concepito in senso ristretto. Per esempio, può adottare campagne educative pubbliche per scoraggiare la violenza sessuale.

In base a certi presupposti, una maggiore intrusione presuntuosa dell'architettura delle scelte da parte dello Stato è particolarmente preoccupante, perché minaccia il benessere, l'autonomia e la dignità. Ma è necessario chiedersi: quali sono quei presupposti, ed è probabile che siano corretti? Perché ed esattamente quando l'ordine spontaneo sarebbe benigno? (C'è qualche forma di darwinismo sociale alla base dell'entusiasmo per l'ordine spontaneo, che rispecchia l'idea che le scelte volontarie e la competizione producono necessariamente buoni risultati? Non è una concezione che si possa difendere facilmente.) È vero che

l’azione del governo crea pericoli e rischi, ma gli sforzi di difendere gli ordini spontanei e le mani invisibili incappano in problemi e obiezioni ben noti, perché non c’è alcun motivo per avere fiducia che gli esiti siano benigni.

Per esempio, un governo che proibisca la discriminazione in base a razza e sesso, o che prenda provvedimenti per garantire un trattamento ragionevole per le persone con disabilità, non si affida affatto a una mano invisibile, e non si fonda su usi e costumi. In ogni caso, l’argomentazione a favore degli ordini spontanei cerca di limitare, soprattutto, il potere coercitivo dello Stato, non i pungoli in quanto tali. Qualunque sia la nostra teoria del campo d’azione legittimo del governo, i danni più gravi tendono ad arrivare da obblighi e proibizioni (da genuina coercizione) e non da pungoli, che conservano la libertà di scelta.

È vero che ordini spontanei, mani invisibili e casualità possono evitare alcuni dei pericoli gravi, e alcuni dei bias caratteristici, che derivano da pungoli presuntuosi da parte del governo. Se temiamo in particolare gli errori dei funzionari pubblici (dovuti a incompetenza o motivazioni negative) vorremo ridurre al minimo le occasioni per i pungoli. E se siamo convinti che i meccanismi della mano invisibile in generale promuovono benessere o libertà, non vorremo disturbare i loro prodotti, anche se fra quei prodotti vi sono dei pungoli.

Secondo me, comunque, la posizione forte a favore degli ordini spontanei e delle mani invisibili non è difendibile. Il mio obiettivo qui, però, non è giustificare quella conclusione. Il punto minimo è che un certo grado di pungolamento da parte dei funzionari dello Stato non è evitabile. Se siamo a favore di ordini spontanei e mani invisibili, saremo a favore di un ruolo specifico per il governo, un ruolo che comprende una specifica architettura delle scelte e pungoli specifici.

L’ASTRAZIONE È UNA TRAPPOLA

Per affrontare le questioni etiche, è importante calare i principi primi dentro le pratiche concrete. A fini di orientamento, sarà utile analizzare più dettagliatamente i pungoli reali e potenziali. Un motivo è evitare la trappola dell’astrazione. Come vedremo, la valutazione etica dei pungoli dipende dal loro contenuto concreto, non dal loro status di pungoli.

I pungoli più ovvi sono costituiti da regole di default, che stabiliscono che cosa succede se le persone non fanno nulla. Altri riguardano la riduzione del pantano e la semplificazione (per esempio, le domande per i permessi di costruzione, la formazione sul posto di lavoro, o gli aiuti finanziari); la divulgazione di informazioni fattuali (come le etichette con l'indicazione delle calorie); avvertimenti, in forma grafica o altra (per esempio, sui pacchetti di sigarette); i promemoria (come quelli di fatture o bollette che stanno per scadere); la maggiore facilità e comodità (per esempio, attraverso il design dei siti web); gli usi di norme sociali (tipo la divulgazione di un confronto fra i propri consumi energetici e quelli dei vicini); ricompense non monetarie, come il riconoscimento sociale; le scelte attive (come nelle domande: "Quale piano pensionistico desideri?" o "Vuoi essere un donatore di organi?"); e le strategie di pre-impegno (attraverso le quali le persone si impegnano, in anticipo, a una particolare linea di condotta, per esempio un programma per smettere di fumare).

È importante notare che alcuni pungoli preservano la libertà di scelta per una popolazione pertinente, mentre impongono un'azione per qualche altra popolazione. Supponiamo, per esempio, che il governo imponga ai grandi datori di lavoro di adottare piani ad adesione automatica per la pensione o per l'assicurazione sanitaria. In questo caso, i dipendenti sono pungolati, ma i datori di lavoro sono costretti. Oppure, supponiamo che il governo imponga agli esercizi di ristorazione (compresi quelli nei cinema) di esibire l'indicazione delle calorie ai clienti. In tal caso i clienti sono pungolati, ma i ristoranti sono obbligati. Alcuni pungoli da parte del governo hanno questa forma: richiedono a un certo gruppo X di pungolare qualche altro gruppo Y. Questa è una cosa da tenere ben presente quando si valutano le obiezioni etiche, perché la coercizione solleva preoccupazioni specifiche, e a volte il governo costringe il settore privato a pungolare i consumatori.

Alcuni pungoli, imposti da enti di regolamentazione, cercano di contrastare il Sistema 1 e di rafforzare la mano del Sistema 2 migliorando il ruolo della deliberazione e i giudizi informati delle persone – per esempio, attraverso la divulgazione di informazioni pertinenti, il *debiasing*[16] e l'uso di strategie di impegno. Altri pungoli sono pensati in modo da attrarre, o attivare, il Sistema 1, come nel caso delle immagini di avvertimento. Alcuni pungoli non fanno appello al Sistema 1, ma funzionano per il modo in cui si comporta – per esempio, nel caso delle regole di default che hanno effetti significativi grazie al potere dell'inerzia.

Un pungolo può essere giustificato in base al fatto che contribuisce a contrastare un bias comportamentale, e (come vedremo) ci sono persone che si oppongono a provvedimenti di questo tipo, specialmente se sembrano mirare o sfruttare il Sistema 1. Un bias comportamentale, però (e questo è un punto importante), *non* è una giustificazione necessaria per un pungolo, e i pungoli non devono necessariamente prendere di mira o sfruttare il Sistema 1 in alcun modo. La divulgazione di informazioni può essere utile anche in assenza di qualsiasi bias. Una regola di default semplifica la vita e pertanto può essere desiderabile, non importa se sia coinvolto o meno un bias comportamentale. Un navigatore GPS è utile anche per le persone che non soffrono di un bias cognitivo che influisce sulla loro capacità di guidare fino alla destinazione desiderata.

Come suggerisce l'esempio del navigatore, molti pungoli hanno l'obiettivo di *aumentare la navigabilità*, cioè di rendere più facile per le persone giungere alla loro destinazione preferita. Simili pungoli hanno origine dall'idea che la vita può essere semplice o difficile da navigare, e un obiettivo di una architettura delle scelte utile è favorire una navigazione più semplice. Finora si è prestata troppo poca attenzione alla relazione stretta fra navigabilità e (buoni) pungoli. Se l'obiettivo è promuovere la navigabilità, le obiezioni etiche risultano molto indebolite e possono addirittura svanire.

I pungoli possono avere un effetto sostanziale sia sulla vita dei singoli, sia sul benessere sociale. In Danimarca, per esempio, l'adesione automatica a piani pensionistici ha avuto un effetto molto maggiore di sostanziosi incentivi fiscali.[17] Negli Stati Uniti, gli sforzi per informare i consumatori sui loro consumi energetici, a confronto con quelli dei loro vicini, ha avuto sui consumi energetici delle famiglie lo stesso (significativo) effetto di un significativo aumento del costo dell'elettricità a breve termine.[18] La semplificazione dei moduli per la richiesta di aiuti finanziari, a favore delle persone che vorrebbero frequentare il college, ha avuto, nel favorire le iscrizioni ai college, un effetto pari a quello di varie migliaia di dollari di aumento negli aiuti finanziari.[19]

VARIETÀ DI PUNGOLI

Al fine di valutare le questioni etiche, sono particolarmente importanti tre distinzioni, perché contribuiscono a chiarire le preoccupazioni di fondo. Innanzitutto, bisogna distinguere i *pungoli paternalistici* dai *pungoli di fallimento del mercato*. Alcuni dei pungoli più familiari sono pensati per proteggere le persone dai loro stessi errori (fra cui anche bias comportamentali); pensiamo a una regola di default concepita per aumentare i risparmi. Altri pungoli sono progettati per rispondere a un tipo comune di fallimento del mercato, che può presentarsi anche se gli individui non compiono alcun errore (e pertanto il paternalismo non è un problema). Pensiamo a una regola di default creata per ridurre l'inquinamento (per esempio, nella forma di una regola di default a favore dell'energia verde).

Nella categoria dei pungoli di fallimento del mercato, possiamo identificare *pungoli per la riduzione delle esternalità*, *pungoli del dilemma del prigioniero* e *pungoli di coordinamento*.[20] Per i pungoli che ricadono in queste categorie, la domanda decisiva dovrebbe essere: aumentano il benessere sociale, correttamente inteso?[21] L'analisi costi-benefici è il migliore fra i modi disponibili per operazionalizzare questa domanda, anche se ha lacune e limiti, non ultimo nel caso in cui si dimostrino rilevanti considerazioni distributive.

C'è ampio consenso sulla legittimità di una risposta, da parte del governo, ai fallimenti del mercato. Se il governo cerca di ridurre un problema di azione collettiva che produce livelli elevati di inquinamento, questo non solleva i tipi di preoccupazioni etiche che entrano in gioco se il governo agisce paternalisticamente. Ne segue che i pungoli di fallimento del mercato non dovrebbero essere particolarmente controversi, in linea di principio, anche se possiamo benissimo porci qualche interrogativo sulla loro efficacia. Di fronte a un fallimento del mercato, un semplice pungolo di solito non è sufficiente; può essere giustificata la coercizione (magari sotto forma di una tassa correttiva, magari nella forma di un obbligo regolativo). Un pungolo però potrebbe dimostrarsi complementare alla coercizione e, in qualche modo, potrebbe essere un suo sostituto.

In secondo luogo, si devono distinguere i *pungoli educativi* da quelli che *non hanno caratteristiche educative, come i pungoli architetturali*. I pungoli educativi cercano di informare le persone, così che possano fare

scelte migliori. Altri, fra cui i pungoli architetturali, sono intesi come ausilio alle persone senza necessariamente aumentarne le conoscenze o la comprensione; le regole di default hanno questa caratteristica. Se l'obiettivo è aumentare la capacità di agire autonomamente delle persone, i pungoli educativi non dovrebbero suscitare particolari obiezioni per ragioni etiche, e in effetti dovrebbero addirittura essere ben accolti – anche se i loro benefici possono non giustificare il loro costo, e anche se possono incontrare problemi di efficacia.

In terzo luogo, si devono distinguere *i pungoli che chiamano in causa o sfruttano bias comportamentali* da quelli che non lo fanno. In particolare, abbiamo visto che alcuni pungoli chiamano in causa o sfruttano il Sistema 1, mentre altri fanno appello al Sistema 2. I tentativi di prendere di mira, o di trarre beneficio da bias comportamentali tendono a essere più controversi, per motivi etici, rispetto a quelli che fanno appello a capacità deliberative. Il motivo è che i primi appaiono più manipolativi e meno rispettosi della capacità di agire autonoma delle persone (un punto su cui tornerò).

Ne segue che i pungoli più controversi sono paternalistici, non educativi e progettati in modo da chiamare in causa o sfruttare bias comportamentali.

RAGIONI ILLECITE E TRASPARENZA

Bisogna riconoscere che l'architettura delle scelte può essere modificata, o che possono venire introdotti nuovi pungoli, per ragioni non lecite. In effetti, molte delle obiezioni più forti ai pungoli, e ai cambiamenti nell'architettura delle scelte, si basano su un giudizio di non liceità delle motivazioni sottostanti. In questi casi, non si fa obiezione ai pungoli in quanto tali, ma alle motivazioni per pungoli particolari.

Per esempio, una regola di default immaginabile potrebbe distorcere il processo democratico dicendo che gli elettori si presume votino per il politico in carica, a meno che non specifichino altrimenti. Una regola del genere violerebbe i principi di neutralità che sono impliciti nelle norme democratiche; per questo motivo sarebbe inaccettabile. In alternativa, un avvertimento potrebbe cercare di spaventare le persone inventando storie su presunti piani nefandi di componenti di un gruppo di minoranza. Si potrebbero invocare norme sociali per spingere le persone ad acquistare

prodotti non sani (“la maggior parte delle persone acquista questi prodotti; anche tu devi farlo!”). In casi estremi, organizzazioni private o pubbliche potrebbero cercare di spingere le persone alla violenza.

Bisogna anche ammettere che la migliore architettura delle scelte spesso richiede una scelta attiva. A volte l’approccio giusto è *imporre* alle persone di scegliere, per assicurarsi che esprimano effettivamente la loro volontà. Altre è meglio *sollecitare* una scelta, chiedendo alle persone che cosa vogliono, senza imporre in alcun modo che lo facciano. Un sollecito è decisamente un pungolo, pensato per fare in modo che le persone esprimano la propria volontà e può non essere accompagnato da alcun tentativo di orientarle in una direzione preferita, ma solo nella direzione di effettuare una scelta.

L’architettura delle scelte dovrebbe essere trasparente e soggetta a scrutinio pubblico, certamente se ne sono responsabili funzionari pubblici. Come minimo, questa affermazione significa che, se quei funzionari istituiscono qualche tipo di riforma, non devono nasconderla al pubblico. Se emanano un regolamento che modifica una regola di default, così da promuovere l’energia pulita o la conservazione, devono dichiarare esplicitamente che cosa fanno, e devono anche spiegare perché lo fanno (facendo riferimento a eventuali bias comportamentali, se quello è il motivo del cambiamento). L’autogoverno stesso richiede uno scrutinio pubblico dei pungoli. Lo scrutinio è un’importante salvaguardia *ex ante* contro i pungoli dannosi; è importante anche come correttivo *ex post*. Trasparenza e scrutinio pubblico possono ridurre la probabilità di una architettura delle scelte che riduca il benessere. Le nazioni inoltre devono trattare i loro cittadini con rispetto, e il consentire uno scrutinio pubblico indica un certo grado di rispetto – e simultaneamente riduce il rischio che i pungoli facciano diminuire il benessere, o violino l’autonomia o la dignità.

Ci si può chiedere se trasparenza e scrutinio pubblico sono sufficienti, anziché puramente necessari. La risposta è che non sono sufficienti. Possiamo immaginare forme di architettura delle scelte che sarebbero inaccettabili anche se fossero pienamente trasparenti; pensiamo a un’architettura (trasparente) progettata per radicare la disuguaglianza sulla base del sesso. Ancora una volta, il problema è che gli obiettivi del pungolo pertinente non sono leciti. Come vedremo, è possibile anche immaginare casi di manipolazione in cui gli obiettivi non sono illeciti, ma in cui il fatto della trasparenza può non essere sufficiente a giustificare un pungolo. Un

pungolo trasparente, annunciato in anticipo ma che assume la forma di una pubblicità subliminale di un tipo o di un altro, andrebbe incontro a obiezioni legittime di manipolazione.

---

1. Alcune delle obiezioni più forti si possono trovare in R. Rebonato, *Taking Liberties: A Critical Examination of Libertarian Paternalism*, Palgrave Macmillan, New York 2012.

2. Vedi R.H. Thaler, C.R. Sunstein, *Nudge. La spinta gentile*, tr. it. Feltrinelli, Milano 2021 ("Per essere considerato un semplice pungolo, l'intervento deve *poter essere evitato facilmente e senza costi eccessivi*", corsivo mio).

3. E. Dayan, M. Bar-Hillel, "Nudge to nobesity II: Menu positions influence food orders", in *Judgment and Decision Making*, 6, 2011, p. 333.

4. Vedi S. Krug, *Don't Make Me Think. Un approccio di buon senso all'usabilità web e mobile*, tr. it. Tecniche Nuove, Milano 2014, p. 11.

5. Vedi B. Keys et al., "Failure to refinance", NBER Working Paper 20798, 2015, pp. 1, 5, https://www.nber.org/papers/w20401.

6. Vedi F.A. von Hayek, "Freedom, reason, and tradition", in *Ethics*, 68, 4, 1958, p. 229. [Una versione leggermente abbreviata di questo saggio si trova, come capitolo 4, con il titolo "Libertà, ragione e tradizione" in F.A. von Hayek, *La società libera*, tr. it. Rubbettino, Soveria Mannelli 2007, pp. 149-181.]

7. E. Ullmann-Margalit, "The invisible hand and the cunning of reason", in *Social Research*, 64, 1997, p. 181, presenta i presupposti in base ai quali le mani invisibili possono essere affidabili, e che cosa metta in dubbio quei presupposti. Per un'ampia analisi del perché le mani invisibili possono produrre danni gravi, vedi G. Akerlof, R. Shiller, *Phishing. L'economia della manipolazione e dell'inganno*, tr. it. Mondadori, Milano 2024.

8. D.F. Wallace, "Questa è l'acqua", in *Questa è l'acqua*, tr. it. Einaudi, Torino 2009, p. 143. [Si tratta della trascrizione del discorso per il conferimento delle lauree tenuto da Wallace al Kenyon College il 21 maggio 2005.]

9. Vedi M.R. Busse et al., "Projection bias in the car and housing markets", NBER Working Paper 18212, 2014, p. 2, https://www.nber.org/papers/w18212.

10. Vedi G. Martin, A. Yurukogly, "Bias in cable news: Real effects and polarization", NBER Working Paper 20798, 2014, https://www.nber.org/papers/w20798.

11. S. Krug, *op. cit.*, pp. 10-19.

12. F.A. von Hayek, *La via della schiavitù*, tr. it. Rusconi, Milano 1995, p. 87.

13. Per un'eccellente analisi delle spiegazioni basate sulla mano invisibile e sui loro limiti, vedi E. Ullmann-Margalit, "Invisible-hand explanations", in *Synthese*, 39, 1978, p. 263.

14. F.A. von Hayek, *La società libera*, cit., p. 166.

15. Vedi in generale F.A. von Hayek, *La società libera*, cit.

16. Vedi in generale C. Jolls, C.R. Sunstein, "Debiasing through law", in *Journal of Legal Studies*, 35, 2006, p. 199.

17. R. Chetty et al., "Active vs. passive decisions and crowdout in retirement savings accounts: Evidence from Denmark", in *Quarterly Journal of Economics*, 120, 2014, p. 1141.

18. Vedi H. Allcott, "Social norms and energy conservation", in *Journal of Public Economics*, 95, 2011, p. 1082.

19. Vedi E. Bettinger et al., "The role of simplification and information in college decisions: Results from the H&R Block FAFSA experience", NBER Working Paper 15361, 2009, https://www.nber.org/papers/w15361.

20. Per il coordinamento, con implicazioni per i pungoli produttivi, vedi E. Ullmann-Margalit, “Coordination norms and social choice”, in *Erkenntnis*, 11, 1977, p. 143.

21. Ovviamente ci sono moltissime discussioni su come vada correttamente inteso il benessere sociale. Per un’analisi preziosa, vedi M.D. Adler, *Well-Being and Fair Distribution*, Oxford University Press, Oxford 2012.

8

PUNGOLI
SENZA MANIPOLAZIONE, 2

Passiamo ora ai pungoli che non sono pensati per affrontare un fallimento comune del mercato e che invece sono intesi come aiuto alle persone perché evitino i propri errori. Pungoli di questo tipo non sono necessariamente manipolativi, ma potrebbero essere considerati paternalistici. Un navigatore GPS non è manipolativo, ma è paternalistico; ti dice di seguire la strada che considera la migliore, dati i tuoi obiettivi. Quando non sono a rischio terze parti, e quando tutto ciò che è coinvolto è il benessere di chi effettua le scelte, l'obiettivo centrale dei pungoli è "influenzare le scelte in modo da migliorare il benessere di coloro che scelgono, *secondo il giudizio di questi ultimi*".[1] In molti casi, quello standard è immediatamente applicabile. Se un navigatore GPS porta le persone a una destinazione che non è quella desiderata, non funziona bene. E se propone loro una strada più lunga e meno comoda, non farà stare meglio le persone, secondo il loro stesso giudizio (a meno che non sia quello che desiderano, magari perché è una strada panoramica).

Molti pungoli possono essere valutati secondo lo standard "secondo il giudizio di chi sceglie"; pensiamo a un promemoria, a un avvertimento, a una regola di default o alla divulgazione di informazioni pertinenti. Per stabilire se lo standard è soddisfatto, dovremmo prendere ciascun pungolo isolatamente. Lo standard però spesso sarà una guida sufficiente.

Purché non ci sia manipolazione, questo modo di intendere l'obiettivo del pungolamento dovrebbe servire a qualcosa per stemperare le principali preoccupazioni etiche. Se l'architetto delle scelte ha effettivamente

successo nel migliorare il benessere di chi sceglie, secondo il giudizio di quest'ultimo, sembrerebbe non esserci alcuna obiezione, dal punto di vista del benessere. E non è chiaro se ci sia un problema dal punto di vista dell'autonomia o della dignità. Tornerò su questi punti.

Allo stesso tempo, bisogna ammettere che lo standard va incontro a interrogativi e problemi seri.

VITE OGGETTIVAMENTE BUONE

Ci sono persone che pensano che la vita di un essere umano possa essere oggettivamente buona o oggettivamente cattiva, e che chi effettua scelte possa compiere, e compia effettivamente, errori oggettivi in merito a ciò che rende buona la sua vita. Come sono intesi nella filosofia politica, gli approcci "perfezionisti"[2] sottolineano l'importanza e la legittimità di aumentare la probabilità che le persone abbiano vite effettivamente buone. Una vita piena di significato, di varietà e di realizzazione può essere considerata oggettivamente buona; si può ritenere invece che una vita dedicata a guardare brevi video su un telefono cellulare non ricada nella stessa categoria. (Ovviamente esistono forme diverse di perfezionismo; qui sorvolo su molte complessità.)

Per le persone che abbracciano il perfezionismo, lo standard del "secondo il giudizio di chi sceglie" si basa su un errore fondamentale: consente che prevalgano i giudizi soggettivi di coloro che effettuano le scelte, anche se sono oggettivamente sbagliati. Immaginiamo, per esempio, che una persona prenda decisioni che le garantiscano una vita breve e non sana, o priva di significato, di varietà o di piacere, o che comporti una grande quantità di sofferenza. Ci si potrebbe chiedere: perché gli architetti delle scelte dovrebbero lasciare mano libera a chi deve scegliere, in circostanze simili?

L'interrogativo solleva questioni gravi all'interno della filosofia politica, a cui non posso dare una risposta qui.[3] Nella misura in cui gli architetti delle scelte lasciano mano libera a chi deve scegliere, può essere che lo facciano per un giudizio morale, secondo il quale chi sceglie ha la sovranità definitiva sulla propria vita, o che lo facciano per umiltà – l'idea che in generale, e nonostante l'esistenza di bias comportamentali, è probabile che abbiano uno svantaggio epistemico rispetto a coloro di cui è in gioco la vita.

Per i nostri scopi qui, il punto fondamentale è che, in quanto adottano lo standard "secondo il giudizio di chi sceglie", gli architetti delle scelte respingono il perfezionismo, e lo fanno non accidentalmente, ma per principio.

PRIMA O DOPO?

Lo standard del "secondo il giudizio di chi sceglie" solleva questo interrogativo: chiediamo quali siano i giudizi di chi deve scegliere prima del pungolo, oppure dopo? I giudizi *ex ante* di chi deve scegliere possono essere differenti dai giudizi *ex post*. Se i giudizi sono influenzati dal pungolo, allora *è possibile che gli architetti delle scelte stiano costruendo proprio il giudizio da cui pretendono di ricavare la loro autorità*. Questo è un problema serio per lo standard del "secondo il giudizio di chi sceglie".

Supponiamo, per esempio, che in base a una regola di default "verde" – che assegni un fornitore di energia attento dal punto di vista ambientale, ma permetta la revoca del consenso e la scelta di un fornitore più economico ma meno rispettoso dell'ambiente – le persone che devono scegliere (o una maggioranza di esse) siano perfettamente soddisfatte. Supponiamo però che quelle persone (o una maggioranza di esse) sarebbero soddisfatte anche con la regola di default opposta. Quali giudizi contano? Ogni volta che il pungolo influisce sui giudizi di chi sceglie, la domanda produce rompicapo gravi.

In qualche caso, è sono scarse le possibilità che il pungolo influisca sui giudizi di chi sceglie: saranno identici, prima e dopo. Quale che sia la regola di default, è improbabile che la maggior parte delle persone non voglia trasformare il 50% del proprio salario in risparmio, o avere un'assicurazione sanitaria che non copre gli interventi chirurgici, o avere pasti che consistono in fette di salame rancido su pane stantio. Quando però i giudizi *ex post* ed *ex ante* sono diversi, diventa più difficile applicare lo standard. Una possibilità sarebbe utilizzare la scelta attiva per vedere che cosa vogliono effettivamente le persone. Un'altra sarebbe valutare il numero delle persone che revocano l'opzione di default sotto regole di default differenti.[4] Una terza possibilità sarebbe tentare un'analisi più diretta del benessere delle persone sotto forme diverse di architettura delle

scelte, anche se bisogna ammettere che qualsiasi analisi del genere fa sorgere difficoltà proprie.

PREFERENZE IN MERITO ALLE PREFERENZE

Un interrogativo distinto nasce dal fatto che le persone non hanno solo preferenze (o preferenze del primo ordine); hanno anche preferenze in merito alle loro preferenze (o preferenze di secondo ordine).[5] Possono voler mangiare cibi deliziosi ma che fanno ingrassare, o spendere ogni mese la maggior parte del loro stipendio, ma potrebbero non voler volere queste cose. Nell'applicare lo standard "secondo il giudizio di chi sceglie", gli architetti delle scelte dovrebbero consultare le preferenze del primo o del secondo ordine?

Alcuni casi immaginabili sono difficili, ma in generale la risposta è immediata: se le preferenze del secondo ordine rispecchiano il modo di pensare del Sistema 2 (inteso, in questo contesto, come i giudizi riflessivi delle persone, opposti ai loro impulsi momentanei), ci sono forti ragioni per sostenere che tali preferenze hanno la precedenza.[6] Certo, il Sistema 1 ha le sue rivendicazioni. Se le persone apprezzano molto certe attività, non è una cosa da trascurare; è parte di ciò che rende la vita degna di essere vissuta. Se un pungolo scoraggia le persone dal mangiare cibi che apprezzano, o dal dedicarsi ad attività che amano, questo significa imporre una genuina riduzione di benessere. Se però le persone pensano, riflettendoci, che il pungolo le spinge in una direzione che approvano, allora lo standard "come giudicato dagli interessati stessi" è soddisfatto.[7]

Quando ci chiediamo quali siano i giudizi di chi deve scegliere, che tipo di informazioni ci aspettiamo che possiedano? Ha senso dire che gli architetti delle scelte dovrebbero appoggiarsi sui giudizi informati delle persone, e non su quelli non informati. Se però gli architetti delle scelte non si concentrano sui giudizi effettivi delle persone che devono scegliere, e si chiedono invece che cosa farebbero quelle persone se fossero informate, c'è il rischio che gli architetti delle scelte si basino sui propri valori e le proprie convinzioni, invece che su quelli delle persone che devono scegliere. In ogni caso, questi architetti potrebbero non avere informazioni sufficienti per sapere se le persone informate reputino di trovarsi in condizioni migliori. Può non essere affatto semplice per chi sta al di fuori mettere a confronto

(dal punto di vista delle persone informate che devono effettuare una scelta) i vari risultati che derivano da pungoli differenti.

In qualche caso, questi punti possono sollevare difficoltà, concettuali ed empiriche, notevoli. Ciononostante, l'idea dei giudizi informati di chi deve scegliere funge da stella polare e impone una vera disciplina. Certo, gli architetti delle scelte dovrebbero concentrarsi sul benessere di chi deve scegliere, e non sul proprio.

Esistono inoltre questioni difficili in merito a come trattare lo standard "secondo il giudizio di chi sceglie" nel caso di problemi di autocontrollo. Supponiamo che qualcuno abbia problemi di quel genere e ne sia consapevole, ma ciononostante desideri, al Tempo 1, di cedere ai suoi impulsi. Dovremmo prendere in considerazione la posizione (1) dell'alcolista, che vuole proprio quella birra; quella (2) del potenziale ex alcolista, che vuole smettere; o quella (3) dell'effettivo ex alcolista, che è grato per essere stato pungolato a smettere di bere? Per certi versi, questa domanda riprende quelle che coinvolgono i giudizi *ex ante* ed *ex post*, e anche quelle che riguardano le preferenze sulle preferenze. Nella misura in cui l'accento però è su problemi di autocontrollo, l'interrogativo è specifico.

È ragionevole sostenere che per la maggior parte gli ex alcolisti sono felici di non essere più alcolisti (perlomeno in assenza di circostanze estremamente inusuali). Per tale motivo abbiamo un forte argomento a favore del fatto che il criterio "secondo il giudizio di chi sceglie" debba essere inteso riferito al giudizio della persona che non è più nella morsa di una dipendenza. Questa affermazione è sufficiente per risolvere molti casi del mondo reale. Ciononostante, può esserci un confine sottile fra un problema di autocontrollo e una legittima attenzione al piacere di breve termine; la questione merita un'analisi più ampia. Nessun architetto delle scelte deve intraprendere un programma di pungolamento che non consideri l'importanza dei piaceri di breve termine, o dei piaceri in generale, che sono, ovviamente, parti fondamentali di una buona vita.

ETICA APPLICATA

Ho già detto che le principali obiezioni etiche ai pungoli riguardano benessere, autonomia e dignità. Mi dedicherò ora a quelle obiezioni. Dirò

qualcosa anche sul pericolo di architetti delle scelte ignoranti o distorti.

L'architettura delle scelte può essere o non essere paternalistica, ma è vero che i pungoli possono essere visti come una forma di paternalismo libertario nella misura in cui cercano di utilizzare l'architettura delle scelte per indirizzare chi deve scegliere verso direzioni che promuovono il suo benessere (ancora una volta, secondo il giudizio di chi sceglie). Se siamo convinti che ogni persona sia il miglior giudice di ciò che le interessa, qualsiasi tentativo di indirizzarla può sembrare malaccorto, perlomeno se l'autonomia o il benessere sono la nostra stella polare.

Per valutare questa obiezione, è importante sottolineare che stiamo parlando qui di una forma particolare di paternalismo, nel senso che è al tempo stesso (a) morbido e (b) orientato ai mezzi. È morbido in quanto evita la coercizione o gli incentivi materiali e quindi preserva a pieno la libertà di scelta. È orientato ai mezzi in quanto non cerca di mettere in dubbio o di modificare i fini delle persone. Come un navigatore GPS, rispetta quei fini (pur con tutte le complessità di cui abbiamo parlato prima). Aiuta le persone ad arrivare dove vogliono andare. Per chi è contrario al paternalismo per motivi legati al benessere, le preoccupazioni più gravi nascono di fronte alla manipolazione o alla coercizione (quando la libertà di scelta è inibita) e quando i pianificatori sociali, o gli architetti delle scelte, non rispettano i fini delle persone. In questo senso, i pungoli aspirano a evitare alcune delle obiezioni welfariste standard al paternalismo.

Ciononostante, alcuni scettici, preoccupati per il benessere, sono contrari al paternalismo in quanto tale.[8] Forse ogni persona è il giudice migliore non solo dei propri fini, ma anche dei mezzi migliori per raggiungerli, dati i suoi gusti e i suoi valori. (Le persone possono non prendere la strada suggerita dal navigatore, perché preferiscono l'alternativa panoramica; il navigatore potrebbe non cogliere facilmente i loro fini, o non servire al loro raggiungimento.) Inoltre, la distinzione fra mezzi e fini non sempre è semplice e immediata. Un problema è il livello di astrazione in cui descriviamo i fini delle persone. Se li descriviamo a un livello di grande specificità (mangiare quel brownie, fumare quella sigaretta, mandare SMS mentre si guida), allora i mezzi delle persone *sono* effettivamente i loro fini. Il brownie è esattamente quello che vogliono; non è un mezzo per qualcos'altro (tranne che per l'esperienza di mangiarlo).

Se invece descriviamo i fini delle persone a un livello elevato di astrazione ("avere una buona vita"), allora quasi tutto è un mezzo per quei

fini. In questo caso, però, non coglieremo le preoccupazioni effettive delle persone; trascureremo ciò che conta per loro. Questi punti sollevano alcuni problemi per quanti sono a favore di una forma di paternalismo orientata esclusivamente ai mezzi. Devono fare attenzione a garantirsi di non descrivere i fini delle persone a un livello di astrazione tanto alto da fraintendere ciò che sta loro a cuore.

Nella misura in cui un navigatore satellitare è un'analogia utile, non è facile considerare discutibili i pungoli per motivi legati al benessere. Molti pungoli sono concentrati completamente sull'aiutare le persone a identificare i mezzi migliori per raggiungere i fini che preferiscono. Consideriamo casi in cui le persone *si sbagliano a proposito dei fatti* (rispetto alle caratteristiche, poniamo, di un prodotto o di un investimento). Se un pungolo le informa, rispetta i loro fini ed è probabile che promuova il loro benessere. Oppure supponiamo che certe caratteristiche di un prodotto siano in qualche modo occultate, e il pungolo aiuti le persone a vederle per quello che sono effettivamente. Oppure supponiamo che le persone soffrano di un bias comportamentale, magari il bias del presente, magari a causa di un ottimismo non realistico. Un pungolo che corregga il loro errore può aiutarle a raggiungere i loro stessi fini.

Certo, alcuni bias comportamentali non sono facili da analizzare in questi termini. Se le persone soffrono di bias del presente, un pungolo è una forma di paternalismo dei mezzi? Supponiamo che le persone mangino cibi molto calorici, oppure bevano parecchio, o non facciano attività fisica, perché pensano a oggi e domani e non tanto all'anno prossimo o al prossimo decennio. Se un pungolo riesce a fare in modo che pensino ai loro interessi di lungo termine, può aumentare il benessere aggregato (intrapersonale) nel tempo. Ma un pungolo del genere è concentrato solo sui mezzi? Se pensiamo che una persona sia una serie di sé che si estendono nel tempo, l'architetto delle scelte a tutti gli effetti ridistribuisce benessere dai sé precedenti a quelli successivi (e magari massimizza anche il benessere delle persone sull'arco della vita). Non è del tutto chiaro, però, se si possa parlare, in questi casi, di paternalismo dei mezzi. E se una persona è considerata un continuo nel tempo, e non una serie di sé, i tentativi di contrastare il bias del presente minano, per ipotesi, i fini di chi deve scegliere, al momento della scelta.

Un motivo fondamentale, per regolatori e altri politici, per rifiutare qualsiasi forma di paternalismo riguarda il benessere: forse le persone sono

i giudici migliori di ciò che promuove i loro interessi, e forse chi è estraneo prenderà abbagli (come credeva John Stuart Mill).[9] Ricordiamo quanto diceva Hayek, che "la consapevolezza della nostra irrimediabile ignoranza della maggior parte di ciò che è noto a qualcuno [che deve compiere una scelta] è *la base principale dell'argomento a favore della libertà*".[10] Una forma di paternalismo che preservi la libertà di scelta, e che sia concentrata sui mezzi, è meno probabile sia deplorevole per motivi legati al benessere, certamente se ci occupiamo di bias comportamentali.

In effetti, è addirittura possibile che i welfaristi alla fine debbano abbracciare il paternalismo coercitivo, almeno di fronte a bias di questo genere.[11] Se il paternalismo migliorasse il benessere, i welfaristi dovrebbero sostenere il paternalismo. Per i welfaristi, il paternalismo dovrebbe essere valutato caso per caso – a meno che non vi sia qualche ragione sistematica per sostenere una posizione contraria al paternalismo.

Forse esiste un buon motivo per tale posizione, radicata nell'idea che chi sceglie con ogni probabilità ha informazioni migliori rispetto agli architetti delle scelte.[12] In qualche caso, però, quell'idea non è corretta, perché chi sceglie non ha conoscenza di tutti i fatti.

I pungoli educativi sono un correttivo naturale. Talvolta è difficile respingere una buona regola di default (poniamo, l'iscrizione automatica a programmi pensionistici) con motivazioni welfariste.

Certo, la scelta attiva può essere migliore di una regola di default, ma questa conclusione non è ovvia; una scelta attiva può imporre oneri reali a chi sceglie (che preferirebbe non scegliere) e potrebbe non ridurre gli errori.[13] I welfaristi possono certo essere inclini a favorire gli approcci che preservano la scelta, in base alla teoria per cui gli individui di solito sanno bene che cosa è più adatto per le loro circostanze, ma, se la nostra guida è il benessere, il fatto che una regola di default abbia una dimensione paternalistica non dovrebbe essere un fattore decisivo contro di essa.

UN MUSCOLO

Esaminiamo qui un altro tipo di idea, che a sua volta nasce da un'attenzione al benessere: *la capacità di scegliere è un muscolo, e la capacità di scegliere bene si può rafforzare con l'esercizio*. Se i pungoli dovessero provocare l'atrofia del muscolo, avremmo un argomento contro

di essi, perché il benessere delle persone dipende dal rafforzamento del muscolo. Potremmo immaginare un'obiezione etica: alcuni pungoli non consentono alle persone di sviluppare le proprie capacità, e possono addirittura indebolire il loro incentivo a farlo. In quel caso, il benessere delle persone si riduce (e dovrebbe essere facile vedere che esiste inoltre un problema di autonomia).

Anche qui è necessario indagare i particolari – il tipo di pungoli e di architettura delle scelte che sono chiamati in causa. La scelta attiva e la scelta sollecitata non impediscono l'apprendimento; non lo fanno neanche le informazioni e i promemoria. Al contrario, promuovono l'apprendimento. I pungoli di questo tipo allenano il muscolo della capacità di scelta, non il contrario. Perlomeno, è così se non sono manipolativi.

Rispetto all'apprendimento, dalle regole di default potrebbe scaturire un problema. È possibile dire che la scelta attiva è di gran lunga meglio dei default, semplicemente perché scegliere può favorire l'apprendimento. Prendiamo, per esempio, la domanda se i datori di lavoro debbano chiedere ai loro dipendenti di compiere scelte attive in merito ai loro piani pensionistici, o se invece debbano di default iscrivere le persone a piani adatti alla loro situazione. Il potenziale di apprendimento può certo andare a favore della scelta attiva; se i dipendenti devono scegliere, e se sono aiutati o incoraggiati ad apprendere, la conoscenza risultante può metterli in una condizione migliore per il resto della loro vita. Se le opzioni di default portano le persone a determinati risultati, non arricchiscono le loro conoscenze, e questa può essere un'opportunità significativa che va persa.

L'argomentazione a favore dell'apprendimento, però, dipende dal contesto. (Ricordate l'analisi precedente dei pungoli educativi.) Esistono campi in cui il muscolo della capacità di scelta deve riposarsi anziché allenarsi. Per la maggior parte delle persone, non è importante diventare esperta nelle numerose decisioni che portano alle impostazioni di default dei telefoni cellulari, perciò l'uso di quelle impostazioni non è esposto a obiezioni. Lo stesso vale in molti altri contesti in cui le organizzazioni si basano sui default invece che sulla scelta attiva. Il risparmio per la pensione può essere uno di questi contesti. Per sapere se gli architetti delle scelte debbano optare per la scelta attiva, è necessario esaminare se il contesto è un contesto in cui, tutto considerato, per chi deve scegliere è prezioso acquisire una certa quantità di conoscenze. Non è necessariamente terribile

se i datori di lavoro creano iscrizioni di default per i piani pensionistici, purché siano nell'interesse di tutti, o della maggior parte dei dipendenti.

È vero, ovviamente, che l'istruzione può essere la risposta appropriata se è probabile che le persone commettano errori (per mancanza di informazioni o per qualche tipo di bias comportamentale). In gran parte dei casi, la prima e la migliore linea di difesa è l'istruzione, che può essere a sua volta caratterizzata come un pungolo, e certamente va considerata una forma di architettura delle scelte. Alcuni, e in particolare lo psicologo Ralph Hertwig, hanno scritto a favore di "cariche" (*boosts*) progettate in modo da migliorare le capacità delle persone; le cariche possono essere preferibili ai pungoli. Abbiamo visto che i pungoli educativi sono una parte importante del repertorio dell'architetto delle scelte. A questo proposito scrive Jeremy Waldron: "Vorrei, però, poter diventare più bravo a scegliere, anziché avere qualcuno sopra di me che sfrutta (anche a mio beneficio) la mia attuale debolezza di pensiero e le mie intuizioni scadenti".[14]

Giusto, ma cariche e istruzione hanno i loro limiti. Le persone ricavano benefici dalle regole di default a proposito di telefoni cellulari, tablet, polizze di assicurazione sanitaria e contratti di noleggio delle auto. Se ognuno dovesse acquisire conoscenze sufficienti su tutto ciò che si riferisce a quegli aspetti, finirebbe rapidamente per esaurire il proprio tempo. La semplificazione spesso è una benedizione. Sul fronte del benessere, una regola di default spesso è l'approccio migliore, perché preserverebbe i risultati desiderabili (ancora, dal punto di vista degli interessati stessi) senza costringere le persone a seguire l'equivalente funzionale di un corso, poniamo, di statistica o di finanza.

Per il welfarista si presenta continuamente una domanda: se, in circostanze particolari, i costi dell'istruzione giustifichino i benefici. Per chi è impegnato in molte attività, sarebbe pretendere troppo insistere sul tipo di istruzione che consentirebbe scelte attive per tutte le caratteristiche pertinenti. Le regole di default possono essere la cosa migliore. Tutto dipende dai fatti, ma, se il benessere è la nostra guida, abbiamo un buon argomento a sostegno della preferibilità delle regole di default sull'istruzione finanziaria, rispetto ad aspetti importanti del trattamento pensionistico.

AUTONOMIA

Lasciamo stare il benessere e torniamo a uno dei nostri temi generali: un motivo indipendente per rifiutare il paternalismo chiama in causa l'autonomia e l'idea del rispetto per le persone. Scrive Stephen Darwall:

> Il carattere discutibile del paternalismo di questo tipo non è che coloro che cercano di darci dei benefici contro i nostri desideri è probabile si sbaglino a proposito di ciò che davvero va a nostro beneficio […]. È, piuttosto, principalmente una mancanza di rispetto, il non riconoscere l'autorità che le persone devono pretendere, di poter fare le proprie scelte da sole.[15]

Mi sembra giusto. I pungoli violano l'autonomia? La risposta dipende da quale sia il tipo di pungolo di cui si parla. Molti pungoli promuovono l'autonomia. Non bisogna considerare autonoma una scelta, se è basata sull'ignoranza; se qualcuno prende un farmaco per il mal di stomaco, convinto che gli sia di aiuto per un raffreddore, non si interferisce con la sua autonomia se lo si informa che quello è un farmaco per il mal di stomaco. L'autonomia richiede scelte informate, e i pungoli educativi sono progettati specificamente per garantire che le scelte siano informate.

A fronte di un bias comportamentale, o di qualche tipo di errore sistematico (secondo il punto di vista dell'attore stesso), non è affatto chiaro che un pungolo violi l'autonomia, correttamente intesa.[16] Quando i pungoli contribuiscono a correggere qualche tipo di bias, possono anche promuovere l'autonomia delle persone, in modo particolarmente ovvio quando aiutano le persone ad avere una migliore comprensione dei fatti. Supponiamo che, a causa dell'euristica della disponibilità, le persone abbiano paura di qualche malattia infettiva (Ebola, poniamo) più di quello che sarebbe giusto date le statistiche, e che un pungolo dia loro un senso più preciso di quanto rischino di contrarre quella malattia. Molti pungoli promuovono la capacità di agire delle persone; non la minano. Un pungolo può promuovere l'autonomia se contrasta il bias del presente o un ottimismo non realistico, oppure consente alle persone di vedere aspetti delle situazioni che altrimenti sarebbero nascosti o invisibili. Come abbiamo visto, ci sono persone preoccupate per la potenziale tensione fra pungoli del governo e capacità individuale di agire. Per i pungoli educativi, però, non esiste una tensione del genere; grazie ai pungoli, in realtà, la capacità di agire autonomamente aumenta.

È importante anche riconoscere che l'autonomia non rende necessarie scelte ovunque. Non giustifica un'insistenza sulla scelta attiva in tutti i contesti. Se dovessimo effettuare delle scelte a proposito di tutto ciò che ha qualcosa a che fare con noi, in poco tempo ne saremmo oppressi. Esiste un rapporto chiaro fra gestione del tempo e autonomia. Alle persone deve essere consentito di dedicare la loro attenzione alle questioni che, dal loro punto di vista, la meritano. Se le persone devono compiere scelte ovunque, la loro autonomia si riduce, se non altro perché non possono concentrarsi su quelle attività che a loro sembrano quelle più degne della loro attenzione.[17] I pungoli architetturali possono promuovere l'autonomia, da questo punto di vista. Danno alle persone più controllo sul tessuto della loro vita. Esempi sono l'iscrizione automatica all'assistenza sanitaria e ai piani pensionistici.

Cionondimeno, è vero che, pensando all'autonomia (e anche al benessere), la migliore architettura delle scelte spesso richiede una scelta attiva. Anche se le regole di default preservano la libertà di scelta, possono violare l'autonomia, perlomeno se non tengono conto delle probabili scelte delle persone. Il problema è che, per la forza dell'inerzia, le persone possono non rifiutare default dannosi. In quel caso, si può ben dire che vi sia una violazione della loro autonomia, perché si ritroveranno con risultati che a loro non piacciono e che non hanno scelto specificamente. Pensiamo, per esempio, a una regola di default in base alla quale, se non indicate altrimenti, si darà per scontata la vostra adesione a un partito politico che invece deplorate, oppure l'intenzione di lasciare la vostra eredità al Vaticano. Anche se le persone possono negare il proprio consenso, le regole di default possono violare l'autonomia in quanto impongono quell'onere alle persone, e in quanto le regole particolari (a) possono rimanere in vigore a causa proprio di quell'onere e (b) non rispecchiano quello che le persone informate preferirebbero.

Per tornare a uno dei miei temi generali: che l'interesse per l'autonomia richieda una scelta attiva, invece del fare affidamento su una regola di default, dipende dalle circostanze. La manipolazione deve essere assolutamente esclusa: un motivo fondamentale è che viola l'autonomia. Rispetto ad alcune dimensioni, le regole di default sono in realtà superiori alla scelta attiva, per ragioni legate all'autonomia. Se le persone scelgono di non scegliere, o se farebbero quella scelta qualora lo si chiedesse loro, è un'offesa alla loro autonomia costringerle a scegliere. Se poi volessero scegliere, una regola di default non le priva di quella possibilità; possono

rifiutare l'opzione di default. Anche nonostante l'inerzia, molte persone faranno così. Preservare la libertà di scelta non è sufficiente ma, purché non sia coinvolta manipolazione, fa molto per garantire che l'autonomia delle persone sia rispettata. Lo stesso si può dire se è richiesto che qualsiasi pungolo paternalistico si concentri sui fini delle persone stesse e per altro abbia obiettivi legittimi. Rispetto all'autonomia, però, continua a esistere il problema della possibilità della manipolazione (su cui torneremo nel capitolo 9).

Se i pungoli rispettano la libertà di scelta, in merito a una dimensione importante preservano l'autonomia delle persone. Quando gli architetti delle scelte costringono le persone, non si limitano più a pungolarle. Se però sono concentrati sull'autonomia, gli scettici possono ancora sostenere che i problemi associati alla coercizione non possono essere totalmente evitati con alcuni pungoli. Abbiamo visto che, per la forza dell'inerzia, le persone possono accettare (passivamente) una regola di default anche se non sono affatto entusiaste del risultato che produce, e rifiuterebbero quel risultato se si concentrassero sul problema relativo. Le regole di default possono rivelarsi manipolative (vedi il capitolo 2).

Dovremmo dubitare che sia corretto descrivere situazioni del genere come situazioni che coinvolgono la coercizione. Nessuno è costretto a fare alcunché. Certo però c'è un rischio che una regola di default produca risultati dannosi. Gli architetti delle scelte devono tener conto di quel rischio e prendere provvedimenti per garantire che la libertà di scelta sia preservata e sia reale. Se è così, la coercizione non è coinvolta.

DIGNITÀ

L'idea della "dignità" è complessa e contestata.[18] Da un certo punto di vista, autonomia e dignità sono in relazione stretta. Sia nella politica che nel diritto, la dignità a volte occupa un posto speciale. Nel diritto costituzionale tedesco, per esempio, la dignità ha un ruolo di primo piano; la Costituzione stessa inizia con queste parole: "La dignità umana è inviolabile [*unantastbar*]. Rispettarla e proteggerla è dovere di ogni potere dello Stato".

Potremmo iniziare dicendo che l'antonimo di autonomia è coercizione, l'antonimo di dignità è umiliazione.[19] Come la intendo qui, l'idea della

dignità richiede rispetto per le persone come agenti; la dignità di una persona non è rispettata se è trattata come un bambino piccolo o una cosa, sottoposta all'autorità superiore di altri. Un intervento non coercitivo, rispettoso dell'autonomia delle persone, può comunque compromettere la dignità. La manipolazione può farlo. Abbiamo visto che, se un medico formula una scelta in un certo modo ("con questo farmaco, può ridurre il suo rischio alla metà"), può manipolarvi, perché quella formulazione rende difficile effettuare una scelta riflessiva e deliberativa. (Supponiamo che il rischio che dovete affrontare sia pari a due su un milione, e il farmaco lo riduca a uno su un milione.) Oppure immaginiamo, per esempio, che un datore di lavoro lasci ai propri dipendenti ampio spazio per la libera scelta, ma parli loro come se fossero bambini che hanno difficoltà a impegnarsi in qualsiasi forma, anche minima, di ragionamento. Un agente di polizia può non trattare i cittadini con dignità anche se rispetta la loro autonomia (sebbene, di nuovo, potremmo intendere l'autonomia in un modo che la rende coestensiva alla dignità). Un agente di polizia può manipolare i cittadini e così violare la loro autonomia. Può sembrare che alcuni pungoli compromettano la dignità e il rispetto per le persone. Possono umiliarle. Come vedremo, questa obiezione è interessante e rilevante, in particolare quando si combina con una preoccupazione per la manipolazione. Si possono immaginare in effetti forme di architettura delle scelte che minano la dignità.

Ovviamente, vi sono interrogativi importanti in merito al posto della dignità nell'etica.[20] Secondo un punto di vista (non convenzionale, va detto), la dignità è propriamente parte di una valutazione del benessere. Se le persone si sentono umiliate, o pensano di essere state trattate senza rispetto, patiscono una perdita di benessere, e quella perdita può essere estremamente grave. In qualsiasi valutazione delle conseguenze per il benessere, una perdita del genere deve essere presa in considerazione. Può darsi che si riveli straordinariamente importante, e vada a deporre contro particolari pungoli.

Per evitare l'ottusità, un buon welfarista dovrebbe riconoscere che un'offesa alla dignità ha un suo carattere qualitativamente particolare. Per sua natura, è un tipo di perdita diverso rispetto alla perdita, poniamo, di denaro o di un'occasione di visitare una spiaggia. Dal punto di vista del welfarista, però, una perdita di dignità è solo un tipo di perdita di benessere, a cui attribuire un peso rispetto agli altri beni che sono in gioco.

Supponiamo, per amor di discussione, che un richiamo evidente e fortemente emotivo, che innesca forti emozioni nel Sistema 1 al fine di scoraggiare le persone dal fumare, sia visto come un'offesa alla dignità, un modo di trattare i fumatori senza rispetto (e magari di infantilizzarli). Qualche fumatore potrebbe percepire in questo modo quel richiamo e contestarlo per quel motivo. Un welfarista però potrebbe essere disposto a sostenere il richiamo emotivo, nonostante la perdita pertinente, se salva un numero significativo di vite.

Da un punto di vista differente, un insulto alla dignità non è semplicemente parte di un calcolo welfarista. Un insulto del genere non dipende dai sentimenti soggettivi delle persone ed è un atto grave, forse specialmente se arriva dalla pubblica amministrazione. Un insulto alla dignità, inteso come mancanza di rispetto nel modo di trattare le persone, non deve essere permesso, a meno che non abbia una giustificazione straordinariamente forte. Se adottiamo questo punto di vista, è particolarmente importante chiedersi se i pungoli offendano la dignità umana.

Per tornare al mio tema generale: la forza dell'obiezione dipende dal particolare pungolo. Un navigatore GPS non insulta la dignità di nessuno. La divulgazione di informazioni fattuali non può certo essere vista come un'offesa alla dignità – meno che mai se le informazioni sono utili e non basate sulla convinzione, falsa e sminuente, che le persone ne abbiano bisogno. Possiamo però immaginare pungoli che offendano l'una o l'altra concezione della dignità, magari perché sono manipolativi. Consideriamo una campagna di salute pubblica, rivolta alla prevenzione dell'obesità, che stigmatizzi e umili le persone sovrappeso, rappresentandole in una luce degradante. O una campagna contro il fumo che faccia la stessa cosa nei confronti dei fumatori.

Qui, di nuovo, il fatto che i pungoli preservino la libertà di scelta e non impongano a nessuno di fare qualcosa, non deve essere preso come una licenza di fare qualsiasi cosa. È possibile immaginare campagne di educazione pubblica che siano manipolative e offendano la dignità, e qualche campagna reale può essere afflitta da questo vizio.

Si può considerare un insulto alla dignità, e una forma di infantilizzazione, anche il fatto che il governo ricordi costantemente alle persone cose che già sanno. Ogni bambino, e tutti quelli che sono stati bambini, possono ricordare questa forma di infantilizzazione, e non sempre

è assente anche nella vita adulta. Se le persone sono informate della stessa cosa ogni ora o ogni giorno (poniamo, dal partner, dal medico o da qualche funzionario pubblico), possono legittimamente percepire che la loro dignità non viene rispettata.

Allo stesso tempo, la maggior parte dei promemoria, degli avvertimenti e degli usi delle norme sociali non sminuiscono nessuno. Né è probabile che gli architetti delle scelte umilino le persone che devono effettuare delle scelte, se diffondono informazioni su quello che fa la maggior parte della gente. Talvolta, la preoccupazione per la dignità può diventare più grave. Se alle persone viene ricordato spesso che incombe una certa scadenza, potrebbero sentirsi trattate come bambini piccoli. Gli avvertimenti possono dare la stessa impressione se sono ripetitivi o condiscendenti, o se sono intesi a innescare forti emozioni, invece di dare semplicemente alle persone un senso di realtà fattuale. In qualche caso, gli avvertimenti possono essere manipolativi.

Anche qui, non c'è obiezione ai pungoli pertinenti in astratto, ma ci possono benissimo essere preoccupazioni per tipi specifici di pungoli. Allo stesso tempo, va sottolineato che l'offesa alla dignità (derivante da promemoria o avvertimenti indesiderati e frequenti) è relativamente minore e, dal punto di vista delle preoccupazioni che hanno prodotto l'attenzione per la dignità nella tradizione politica occidentale, è ridicolmente modesta.

Qual è il rapporto fra dignità e regole di default? Se un datore di lavoro iscrive automaticamente i suoi dipendenti a piani pensionistici e sanitari, non è certo a rischio la dignità. Se un'azienda che commercializza telefoni cellulari adotta una serie di impostazioni predefinite per il telefono e di regole di default per le condizioni contrattuali, non va perso nulla in termini di dignità. Possiamo immaginare, però, casi più difficili. Supponiamo che il governo richieda "pasti di default" in vari ristoranti, in modo che alle persone vengano presentate scelte sane, a meno che non scelgano specificamente altro. Mettiamo da parte il fatto che, rispetto ai ristoranti, questo approccio è un obbligo e non un semplice pungolo. Una risposta plausibile si basa su una preoccupazione per la dignità: perché pretendere pasti di default? Perché non chiedere semplicemente alle persone di scegliere liberamente quello che vogliono? Oppure supponiamo che un governo specifichi un "piano di attività fisica di default" per gli adulti, così che si presuma che vogliano dedicarsi a certe attività, a meno che non vi

rinuncino esplicitamente. Le persone potrebbero dare la stessa risposta plausibile, magari con una notevole mobilitazione.

Va notato che le regole di default di questo tipo possono essere detestabili sia per i welfaristi che per i non welfaristi. I primi potrebbero volersi concentrare sui sentimenti soggettivi delle persone. Se le persone pensano di essere trattate come bambini, e se si offendono per quel trattamento, i welfaristi dovrebbero tenerne conto nella loro valutazione di costi e benefici. I non welfaristi insisterebbero che l'offesa alla dignità è da respingere anche se ha qualche tipo di giustificazione welfarista.

In situazioni estreme, le regole di default potrebbero effettivamente costituire un affronto grave alla dignità. In quel caso, dovrebbe esserci una forte presunzione contro di esse (quali che siano le nostre posizioni fondamentali). Ancora una volta, però, sarebbe un errore utilizzare situazioni estreme o casi a malapena immaginabili per motivare una opposizione alle regole di default in generale. Le persone non vengono trattate senza rispetto se un'istituzione adotta come opzione di default la stampa fronte-retro o se vengono automaticamente iscritte a piani pensionistici o assicurazioni sanitarie. L'obiezione sulla base della dignità ha molta più forza in astratto che nel contesto di tutti, o quasi tutti, i casi reali in cui sono in vigore effettivamente delle regole di default.

FUNZIONARI TENDENZIOSI

Gli architetti delle scelte sono decisamente esseri umani, pienamente soggetti a bias comportamentali; spesso sono inaffidabili. Il settore in crescita della scelta pubblica comportamentale sviluppa questo punto per offrire un'analisi dell'errore dei funzionari.[21] È ragionevole fare obiezione ad alcuni pungoli, e ad alcuni tentativi di intervenire nell'architettura delle scelte esistente, sulla base del fatto che gli architetti delle scelte possono prendere abbagli. Possono non disporre di informazioni importanti (problema della conoscenza). Possono essere vittima di distorsioni, magari perché sono in gioco i loro interessi particolari (problema della scelta pubblica). Possono presentare loro stessi dei bias comportamentali, come il bias del presente, il bias ottimistico o il *probability neglect*. In una società democratica, i funzionari pubblici sono attenti all'opinione pubblica e, se il pubblico è in errore, possono sbagliare anche i funzionari.

Non è chiaro se e fino a che punto questa obiezione sia un'obiezione chiaramente etica, ma identifica un aspetto su cui è importante essere cauti. Un motivo per i pungoli, in contrapposizione a obblighi e proibizioni, è che gli architetti delle scelte possono sbagliare. È un'affermazione su cui nessuno dovrebbe essere in disaccordo, e va fortemente a favore di approcci che preservino la scelta. Se gli architetti delle scelte prendono abbagli, almeno si può dire che le persone hanno il diritto di andare per la loro strada. Se mettiamo in evidenza il rischio di errori da parte dei pubblici funzionari, potremmo voler evitare anche pungoli e architettura delle scelte.

La risposta iniziale a questa obiezione dovrebbe essere familiare: un'architettura delle scelte è inevitabile. Quando gli architetti delle scelte agiscono, modificano l'architettura; non creano un'architettura dove prima non ne esisteva alcuna. Un certo grado di pungolamento dal settore pubblico non può essere evitato, e non ha alcun senso desiderare la sua scomparsa. Ciononostante, gli architetti delle scelte che lavorano per il governo potrebbero decidere che è meglio cercare di basarsi sui mercati liberi e confidare nei mercati liberi e nei meccanismi della mano invisibile. In tal caso, opterebbero per (o accetterebbero) un'architettura delle scelte che rispecchi quei meccanismi.

Questa idea presenta molti rompicapo concettuali ed empirici, a cui ho fatto cenno in precedenza, e che non approfondirò ulteriormente qui. Quello che ci si chiede è se sia tanto astratta, e tanto radicata nel dogma, da non poter pretendere alcun sostegno. Certo, i mercati liberi hanno le loro virtù, ma in qualche caso divulgazione, avvertimenti e promemoria possono fare di gran lunga più bene che male. Come abbiamo visto, la scelta attiva a volte è inferiore alle regole di default. Qualcuno deve decidere a favore dell'una o delle altre e, in qualche caso, quel qualcuno è inevitabilmente il governo. È vero che la sfiducia nei confronti dei funzionari pubblici può essere una motivazione contro i pungoli, almeno quando sono evitabili, ma se dogmatica e generalizzata quella sfiducia con tutta probabilità produrrà gravi perdite in termini di benessere, autonomia e addirittura dignità. Le mani invisibili possono compromettere tutte quelle cose, e lo fanno.

LA CAPACITÀ DI AGIRE COME VALORE PRINCIPE

È inutile rifiutare pungoli e architettura delle scelte in sé. Gli esseri umani non possono vivere in un mondo che ne sia privo. Gli ordini spontanei e le mani invisibili hanno molti pregi, ma pungolano. Anche un governo ridotto alle minime dimensioni deve creare architetture delle scelte di molti tipi diversi. Lo stesso diritto consuetudinario influisce sulle decisioni delle persone, che sia stato progettato per farlo o meno (o che sia stato progettato o meno). Torniamo agli effetti delle regole di default, di quel tipo onnipresente nel diritto civile di contratti, proprietà e responsabilità civile.

Molti pungoli e molti cambiamenti nell'architettura delle scelte non sono semplicemente ammissibili su basi etiche; sono di fatto indispensabili. Per quanto riguarda il benessere, la considerazione dovrebbe essere immediata: gran parte dei pungoli promettono di aumentare il benessere sociale. Questo però vale anche per autonomia e dignità. In qualche caso esiste una tensione immaginabile fra pungoli e capacità umana di agire. Che i pungoli compromettano quella capacità di agire, però, dipende dal loro contenuto, non dal loro status di pungoli. Come abbiamo visto, molti pungoli promuovono la capacità di agire.

È importante avere un senso abbastanza ampio della categoria dei pungoli, e una conoscenza piena delle differenze fra di essi. Per fare dei passi avanti, le preoccupazioni etiche devono essere calate negli esempi particolari. È vero che qualsiasi modifica nell'architettura delle scelte, comprese quelle che preservano la libertà, può incorrere in obiezioni etiche gravi e addirittura convincenti – il caso più ovvio è quello in cui gli obiettivi di fondo sono illeciti. Se gli obiettivi sono legittimi, però, è meno probabile che i pungoli vadano incontro a vincoli etici, non ultimo quando promuovono scelte informate (come nel caso dei promemoria). Trasparenza e scrutinio pubblico sono salvaguardie importanti, specialmente quando responsabili di pungoli e architettura delle scelte sono funzionari pubblici. Non deve esserci nulla di nascosto o di occultato.

La storia delle nazioni che rispettano la libertà è piena di cambiamenti nell'architettura delle scelte che sono stati motivati da un'aspirazione a realizzare gli ideali nazionali più elevati e più duraturi. Nell'avvicinarsi a quegli ideali, si dimostreranno indispensabili nuovi pungoli, che rafforzino la capacità individuale di agire.

1. R.H. Thaler, C.R. Sunstein, *Nudge. La spinta gentile*, tr. it. Feltrinelli, Milano 2022, p. 11.

2. Una forma di perfezionismo liberale è difesa in J. Raz, *The Morality of Freedom*, Clarendon Press, Oxford 1986. Una critica del perfezionismo si trova in S. Conly, *Against Autonomy*, Cambridge University Press, Cambridge (UK) 2013, pp. 100-125.

3. Analisi pertinenti si possono trovare in S. Conly, *op. cit.*; B.D. Bernheim, "The Good, the Bad, and the Ugly: A unified approach to behavioral welfare economics", in *Journal of Benefit-Cost Analysis*, 71, 2016, p. 12; M. Nussbaum, *Creare capacità*, tr. it. il Mulino, Bologna 2012; A. Sen, *Commodities and Capabilities*, Oxford University Press, Oxford 1999. Per un breve, brillante insieme di obiezioni al perfezionismo, vedi S. Conly, *op. cit.*, pp. 109-125.

4. R.H. Thaler, C.R. Sunstein, *op. cit.*

5. Vedi H. Frankfurt, "Freedom of the will and the concept of a person", in *Journal of Philosophy*, 68, 1971, p. 5.

6. Qui sorvolo su qualche complessità. Per esempio, le intuizioni delle persone possono essere ben addestrate, e i loro giudizi riflessivi possono essere inaffidabili. Possiamo immaginare un tennista che sa esattamente che cosa fare quando il primo servizio dell'avversario va al suo rovescio, anche se non se la caverebbe bene a spiegare riflessivamente, a se stesso o ad altri, che cosa fare in quelle circostanze.

7. Si presentano, però, alcune domande difficili sollevate dall'incoerenza temporale. Supponiamo che al Tempo 1 Edward voglia comportarsi in un certo modo, lo faccia e ne ricavi un guadagno in termini di benessere, ma, al Tempo 2, Edward desiderasse che il suo sé precedente non si fosse comportato in quel modo, e sperimenti una perdita di benessere. In tal caso, lo standard "secondo il giudizio di chi sceglie" finisce per incappare in difficoltà genuine. Vedi D. Whitman, M. Rizzo, "The problematic welfare standards of behavioral paternalism", in *Review of Philosophy & Psychology*, 6, 3, 2015, pp. 409-425.

8. Vedi, per esempio, R. Rebonato, *Taking Liberties: A Critical Examination of Libertarian Paternalism*, Palgrave Macmillan, New York 2012; J.D. Wright, D.H. Ginsburg, "Behavioral law and economics: Its origins, fatal flaws, and implications for liberty", in *Northwestern University Law Review*, 106, 2012, p. 1033.

9. J.S. Mill, *Sulla libertà*, tr. it. Bompiani, Milano 2000.

10. F.A. von Hayek, *The Market and Other Orders*, in B. Caldewell (a cura di), *The Collected Works of F.A. Hayek*, The University of Chicago Press, Chicago 2013, p. 384 (corsivo mio).

11. S. Conly, *op. cit.*, p. 32.

12. *Ibidem*, p. 211. Per un'altra concezione non welfarista di che cosa c'è di sbagliato nel paternalismo, vedi N. Cornell, "A third theory of paternalism", in *Michigan Law Review*, 113, 2015, p. 1295 (dove si sostiene che il paternalismo può essere sbagliato perché implica, ed esprime un giudizio, che chi sceglie non sia in grado di compiere buone scelte per se stesso).

13. Vedi C.R. Sunstein, *Choosing Not to Choose: Understanding the Value of Choice*, Oxford University Press, Oxford 2015, pp. 144-149.

14. Vedi J. Waldron, "It's all for your own good", in *New York Review of Books*, 2014, https://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good. Sui boost, vedi R. Hertwig, "When to consider boosting: Some rules for policy-makers", in *Behavioral Public Policy*, 1, 2017, p. 143.

15. Vedi S. Darwall, "The value of autonomy and the autonomy of the will", in *Ethics*, 116, 2006, pp. 263, 268.

16. Vedi S. Conly, *op. cit.*, p. 36: "Anche se ammettiamo che gli individui hanno diritti, e quindi possono pretendere di non essere danneggiati da altri in certi modi, o che altri ancora li difendano in queste pretese, perché dovrebbe esistere qui un diritto del genere, dato che il senso dell'azione è aiutare la persona a raggiungere quello che, sul lungo periodo, vuole, e quello che vorrebbe se non fosse un pensatore fallace?".

17. Vedi S. Mullainathan, E. Shafir, *op. cit.*, p. 46.

18. Per un’analisi particolarmente preziosa, vedi M. Rosen, *Dignità: storia e significato*, tr. it. Codice, Torino 2013. Nel contesto legale, vedi C. McCrudden, “Human dignity and judicial interpretation of human rights”, in *European Journal of International Law*, 19, 2008, p. 655.

19. Vedi A. Margalit, *La società decente*, tr. it. Guerini, Milano 1998, pp. 57-72.

20. *Ibidem*. Vedi in generale M. Rosen, *op. cit.*; C. Beitz, “Human dignity in the theory of human rights”, in *Philosophy & Public Affairs*, 41, 2013, p. 259; T. Christiano, *Two Conceptions of Human Dignity as Persons*, manoscritto non pubblicato, 2008.

21. Per un esempio, vedi T. Kuran, C.R. Sunstein, “Availability cascades and risk regulation”, in *Stanford Law Review*, 51, 1999, p. 683. Per una panoramica, vedi J. Schnellenbach, C. Schubert, “Behavioral public choice: A survey”, 2014, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2390290.

9

INTELLIGENZA ARTIFICIALE: PROMESSA E MINACCIA

Racconti di un futuro distopico spesso presentano due forme diverse di schiavitù umana. La prima fa perno sulla paura del dolore; la seconda sul fascino del piacere. I manipolatori vanno a nozze sia con il dolore, sia con il piacere.

Nel XX secolo, le due forme di schiavitù sono state definite, rispettivamente, da *1984* di George Orwell e da *Il mondo nuovo* di Aldous Huxley. Il romanzo di Orwell è l'incubo di uno Stato totalitario, in cui il Grande Fratello governa con il terrore e in cui chiunque è potenzialmente esposto al carcere o alla tortura. Poiché di Grandi Fratelli se ne possono trovare ovunque nel mondo, e per il racconto feroce di come si possa spezzare la libera volontà ("Amava Grande Fratello"), il libro di Orwell continua ad avere risonanza.

Nelle parole di Orwell, arresti e terrore sono accompagnati da una dose abbondante di manipolazione. Considerate: "Potere vuol dire ridurre la mente altrui in pezzi che poi rimetteremo insieme nella forma che più ci parrà opportuna"; oppure: "È qualcosa di bello, la distruzione delle parole"; o ancora, soprattutto: "Noi non ci limitiamo a distruggere i nostri nemici, noi li cambiamo".[1]

Il libro di Huxley, invece, è un incubo di consumismo, in cui le persone sono schiave del desiderio di piacere, fornito soprattutto dal "soma", una droga fittizia. Nel mondo di Huxley, le persone si divertono molto, ma la loro vita non ha significato e non esistono connessioni genuine. Poiché quel mondo è al tempo stesso seducente e distruttivo per l'anima, e poiché il

libro suscita un senso di riconoscimento nelle società libere, anch'esso continua a dirci qualcosa. C'è molta manipolazione nel mondo di Huxley. Per esempio: "Si credono le cose perché si è stati condizionati a crederle"; oppure: "Se uno è diverso, è fatale che sia solo".[2]

Né Orwell né Huxley avevano molto da dire sulla personalizzazione, che invece è una caratteristica tipica della nostra epoca, e che aumenta significativamente le occasioni per manipolare le persone. Né avevano qualcosa da dire su algoritmi o intelligenza artificiale, che possono essere altri motori di manipolazione. Amazon.com sa quali libri e quali film avete acquistato, e vi offre suggerimenti sulla base di quegli acquisti. Sa che cosa vi piace. Dite qual è il vostro cantante o il vostro brano preferito, e l'intelligenza artificiale può crearvi una stazione radio solo per voi. Dopo aver tenuto traccia dei vostri movimenti nel Web, Google sa molte cose su ciò che vi interessa, e mette in evidenza dei prodotti sulla base di quello che sa. Siri sa che cosa avete in agenda, e sa molto dei vostri gusti e delle vostre preferenze. Cambridge Analytica sperava di acquisire un'enorme quantità di dati su un numero enorme di persone, e di usare quei dati per spostare le persone nelle direzioni desiderate. Le grandi aziende hanno sempre più potere per farlo. E lo stesso vale per i governi. Tutto questo mette il problema della manipolazione in netto rilievo, in un modo che non ha precedenti.

LEI

Lei, lo splendido film del 2013 diretto da Spike Jonze, rappresenta un mondo distopico di piacere, ma in una forma tipicamente contemporanea. Theodore Twombly, il protagonista del film, si guadagna da vivere scrivendo lettere e cartoline molto personalizzate – per esempio, biglietti di auguri dalle mogli ai mariti per l'anniversario di matrimonio. Lo fa dopo aver raccolto molte informazioni sia sul mittente, sia sul destinatario. Nel mondo di Twombly, le lettere d'amore sono al tempo stesso esternalizzate e personalizzate. Twombly non è esattamente un sistema operativo, ma si comporta in maniera molto simile. Nel suo modo particolare, è un manipolatore.

Deve anche affrontare un imminente divorzio. La sua vita va a rotoli, piena di videogiochi e di sesso telefonico anonimo. Tutto cambia quando

acquista un sistema operativo, una forma di intelligenza artificiale (pensate a una Siri 4.0), che si chiama Samantha.

Samantha ha accesso al computer di Twombly, e anche alle sue e-mail. Sa leggere molto rapidamente. Sa che cosa gli piace, che cosa no. Conosce i suoi punti di forza e i suoi punti deboli. Sa che cosa lo entusiasma. Sa che cosa lo deprime.

Soprattutto, forse, sembra interessata a lui. Ascolta. Può vedere il mondo attraverso i suoi occhi. È lì quando si sveglia, ogni notte lei è quella che gli augura la buona notte e quella a cui la augura lui. Lo veglia mentre dorme.

Se fosse tutto qui, ovviamente, l'interesse di Twombly potrebbe sfumare rapidamente. Se non sei un narcisista impenitente, non puoi innamorarti di qualcuno che sa dire soltanto "Dimmi di più". Per come è costruita, Samantha ha interessi e preoccupazioni proprie. Le piace scrivere musica, ama giocare, è curiosa, può essere insicura ed è un po' sbarazzina. Non lo sappiamo per certo, ma forse anche quelle caratteristiche sono un prodotto della personalizzazione. Forse sono esattamente quello che Twombly vuole e di cui ha bisogno. L'algoritmo di Samantha sembra in grado di identificare quelle cose. Samantha può di manipolare Twombly come le pare.

Qualcuno parla di "innamorarsi" del proprio nuovo telefono o tablet. È un modo di dire, ma Twombly invece si innamora proprio, in pieno e con gioia, di Samantha (incredibilmente, Jonze riesce a far funzionare la vicenda). *Lei* sarebbe stato un film curioso ma meno valido (un aggiornamento più brillante e più felice de *Il mondo nuovo*), se fosse finito qui. Nelle relazioni umane, però, le persone cambiano e, quando succede, quelle che un tempo andavano perfettamente d'accordo possono finire per allontanarsi. Samantha è un sistema operativo, ma anche lei cambia, lascia Twombly e gli spezza il cuore. Uno dei temi del film è quanto ci sia simultaneamente di conforto, stimolazione e artificialità nei legami con ciò che non è umano – telefoni, tablet, sistemi operativi di ogni genere. La loro attrattiva sta, fra l'altro, nella loro capacità di imparare, quasi immediatamente, di che cosa hanno bisogno le persone e che cosa piace loro, e di manipolarle di conseguenza, soprattutto catturando la loro attenzione e attraendo anche i loro interessi finanziari.

Lei, ovviamente, non è l'unica rappresentazione fantascientifica del potere che ha potenzialmente l'intelligenza artificiale di manipolare gli esseri umani. Una versione molto più cupa si trova in *Ex Machina*, un film

del 2014. Parla di Ava, non un sistema operativo ma un (bellissimo) robot umanoide, costruito da un progettista brillante, Nathan Bateman, con l'aiuto di Caleb Smith, un programmatore.

Ava e Caleb formano quella che sembra una connessione, che diventa qualcosa di simile all'amore, perlomeno da parte di Caleb. Per molto tempo non sappiamo se Ava sia motivata da qualcosa, ed eventualmente da cosa: davvero si innamora di Caleb? In ogni caso, è abile a toccare tutti i punti giusti e a farlo innamorare di lei. Si dà il caso che Ava sia una manipolatrice eccezionale. È interessata alla propria libertà, non all'amore. Riesce a ottenere da Caleb quello che vuole da lui, che è la libertà. Porta sia Caleb che Nathan alla morte.

MOTORI DI SCELTA

I motori di scelta[3] possono essere definiti meccanismi, in genere (ma non sempre) online, attraverso i quali a chi deve effettuare scelte viene data la possibilità di fornire alcune informazioni su di sé e sulle proprie preferenze, e quindi ricevere una raccomandazione, o un insieme di raccomandazioni, in merito a quello che dovrebbe scegliere. Supponiamo che dobbiate scegliere un nuovo laptop. Potete dire che volete pagare meno di 1200 euro; che volete un modello che pesi meno di un chilo e mezzo; e che abbia una batteria che garantisca un'autonomia di almeno nove ore. Un motore di scelta può darvi una serie di opzioni, utilizzando l'intelligenza artificiale o anche no. Potrebbe essere un algoritmo: le persone con le caratteristiche e le preferenze A, B, C in genere preferiscono, nell'ordine, i laptop X, Y, Z. Qualsiasi tipo di motore di scelta basato sull'intelligenza artificiale potrebbe specificare e restringere il ventaglio di opzioni.

Motori di scelta di questo genere si potrebbero utilizzare per qualsiasi scelta: tanto per fare qualche esempio a caso, per scegliere cani,[4] libri gialli, cellulari, rasoi per la barba, partner romantici, scarpe, racchette da tennis o cravatte. I motori di scelta possono usare o meno l'intelligenza artificiale; se la usano, può essere intelligenza artificiale di vario tipo.

Un modello linguistico di grandi dimensioni (LLM, *Large Language Model*) può fungere da motore di scelta; può presentare delle opzioni quasi istantaneamente. Nel 2024, ho dato a ChatGPT questo prompt: "Vivo in una

casa in periferia. Mi piacciono i cani grandi. Quale razza di cane dovrei scegliere?”.

Ecco la risposta che ho ricevuto:

> Vivere in periferia dà il vantaggio di più spazio, che è un bene per le razze di maggiori dimensioni. Ecco alcune razze di cani grandi che potrebbero andare bene:
>
> 1. **Golden retriever:** amichevole, paziente, ben noto per il comportamento tranquillo. Richiede attività fisica regolare, ma in generale è facile da addestrare.
> 2. **Labrador retriever**: pieno di energia, socievole e amichevole. I Labrador vanno molto bene per le famiglie e amano essere attivi.
> 3. **Pastore tedesco**: intelligente, fedele e protettivo. È facilmente addestrabile ed è un compagno eccellente.
> 4. **Bovaro del Bernese**: dolce, affettuoso, di buona indole. In generale tranquillo, ma ha bisogno di esercizio fisico regolare.
> 5. **Alano**: nonostante le dimensioni, questi cani sono famosi per essere giganti gentili. Sono amichevoli, affettuosi, ottimi per la famiglia.
> 6. **Mastino**: fedele, protettivo, di buona indole. Sono cani relativamente poco pieni di energia per la loro taglia, ma hanno bisogno di spazio per le loro grandi dimensioni.
> 7. **Terranova**: noto per il temperamento dolce e per l’eccellente abilità nel nuoto. Sono cani dolci, pazienti, buoni con i bambini.
> 8. **Rottweiler**: sicuro, fedele e protettivo. Con l’addestramento e l’opportuna socializzazione, sono ottimi animali per le famiglie.
>
> Assicurati di tenere conto del livello di attività, delle esigenze di cura e del temperamento di ciascuna razza, per trovare quella che corrisponde meglio al tuo stile di vita. Inoltre, visite ai canili e rifugi locali possono offrire opportunità di trovare un cane di grossa taglia che ha bisogno di una casa.

È una risposta eccellente. Conosco abbastanza bene i cani, e la risposta corrisponde quasi perfettamente ai miei gusti (anche se la mia preferenza va ai Labrador). Non è manipolativa, anche se ha qualche formulazione entusiastica (ma giustificabile).

Quello che mi interessa qui è come i motori di scelta, basati sull’intelligenza artificiale e autorizzati o richiesti dalla legge, possano ovviare a un’assenza di informazione o a un bias comportamentale, e quindi possano aumentare il benessere sociale. Nonostante *Lei* ed *Ex Machina*, voglio attirare l’attenzione sull’importanza della personalizzazione e sulla promessa di motori di ricerca basati sull’intelligenza artificiale, per migliorare le scelte individuali. Sarà un esame ampiamente ottimistico, concentrato sulla promessa straordinaria dei motori di scelta. Alla fine, però, mi concentrerò anche sulle minacce, compresi i rischi di inganno e manipolazione.

PIANIFICAZIONE E BIAS

Per i piani pensionistici, molti datori di lavoro utilizzano qualcosa di simile a un motore di scelta – una versione rozza, non molto seducente, di Samantha e Ava. I motori di scelta sanno un po' di cose (magari anche tante) dei loro dipendenti e, sulla base di ciò che sanno, iscrivono automaticamente le persone a un piano specifico. Il loro scopo è aiutare, non manipolare.

Il piano scelto spesso è un fondo indicizzato diversificato e gestito passivamente. I dipendenti possono rifiutarlo e scegliere un piano diverso, se lo desiderano. In alternativa, i datori di lavoro possono offrire ai dipendenti un insieme specificato di opzioni, nel presupposto che siano tutte adatte, o abbastanza adatte. Le opzioni non adatte di solito non vengono incluse. I motori di scelta possono dare alle persone informazioni semplici per scegliere fra le opzioni, che possono essere identificate o ampliate con l'assistenza dell'intelligenza artificiale o di qualche tipo di algoritmo. Se gli utenti le chiedono, possono ricevere più informazioni e informazioni più complesse.

Ecco un'impostazione ragionevole: *iscrivere automaticamente i dipendenti a un piano che ha più probabilità di migliorare il loro benessere, dato tutto ciò di pertinente che è noto su di loro*. L'identificazione di quel piano può dimostrarsi molto impegnativa, ma come minimo si può escludere un gran numero di piani. Va notato che, se ci si concentra sul migliorare il benessere delle persone, non si parla necessariamente di preferenze rivelate. Si parla di ciò che migliorerebbe realmente il benessere dei dipendenti, non necessariamente di ciò che sceglierebbero (anche se possiamo sperare che le due cose siano molto simili, o magari in ultima istanza siano la stessa cosa).

Per quanto riguarda il risparmio pensionistico, possiamo facilmente immaginare tanti tipi diversi di motori di scelta. Alcuni possono essere manipolativi; alcuni elaborati; alcuni casuali; alcuni grezzi o incapaci; alcuni possono mostrare bias comportamentali o altri bias; alcuni possono essere interessati. Per esempio, le persone potrebbero essere iscritte automaticamente a piani con canoni ingiustificatamente elevati. Potrebbero essere iscritte automaticamente a piani non diversificati. Potrebbero essere iscritte automaticamente a conti di deposito del mercato monetario, senza alcuna speranza di rivalutazione. Potrebbero essere iscritte automaticamente

a piani dominati (cioè a piani che sono chiaramente peggiori di altri). Potrebbero essere iscritte automaticamente a piani che sono decisamente inadatti alla loro situazione. Potrebbero vedersi offrire un gran numero di opzioni, per poi dover scegliere con poche informazioni pertinenti o con informazioni complesse che le inducono a fare scelte poco oculate. Quest'ultimo caso può essere considerato una forma di manipolazione.[5]

Quello che mi interessa qui, però, non sono i piani pensionistici. La considerazione generale è che, in linea di principio, i motori di scelta potrebbero essere utili per compensare una carenza di informazioni o bias comportamentali.[6] Potrebbero essere una sorta di Samantha, non manipolativa e non interessata a relazioni romantiche, ma attenta ad aiutare le persone e molto brava nel fare proprio quello. Per i piani pensionistici, i motori di scelta potrebbero essere paternalistici, oppure non esserlo. Se non lo sono, forse è perché offrono semplicemente un menu di opzioni, in base a quello che sanno delle persone che devono fare una scelta. Se sono paternalistici, potrebbero esserlo poco, moderatamente o molto. Un motore di scelta moderatamente paternalistico potrebbe creare barriere non banali a quanti cercano certi tipi di piani (per esempio, quelli con canoni elevati). Le barriere possono assumere la forma di fornitura di informazioni, domande del tipo "sei sicuro di volerlo?" e obbligo di molti clic (pantano!). Possiamo pensare che un motore di scelta moderatamente paternalistico offra "pattern chiari", anziché pattern oscuri. Un motore di scelta molto paternalistico potrebbe proibire ai dipendenti di selezionare qualsiasi piano diverso da quello che il motore stesso ritiene sia nel loro interesse, oppure potrebbe rendere loro estremamente difficile sceglierlo.

Motori di scelta di questo tipo potrebbero essere utilizzati per un numero qualsiasi di scelte. Potrebbero usare l'intelligenza artificiale e, nel caso, intelligenza artificiale di tipi diversi. Un LLM potrebbe funzionare come un motore di scelta; potrebbe offrire opzioni quasi istantaneamente. Prendiamo questa domanda: quale modello di automobile vorreste acquistare? Un'auto che vi costi 800 euro più dell'alternativa, ma che vi faccia risparmiare 8000 euro nell'arco dei prossimi dieci anni? Acquistereste un frigorifero ad alta efficienza energetica che vi costerebbe *x* oggi, ma vi farebbe risparmiare dieci volte tanto nell'arco dei prossimi dieci anni? Quali caratteristiche di un'auto o di un frigorifero contano di più per voi? Avete bisogno di una macchina grande? Vi piacciono le ibride? Le elettriche vi attirano, oppure non tanto?

Come abbiamo visto, i consumatori sono affetti dal “bias del presente”.[7] I costi e i benefici di oggi pesano moltissimo; i costi e i benefici futuri no. Per molti di noi, quello che conta di più è il breve termine; il lungo termine è un altro mondo. Il futuro è Futurlandia, un paese che non siamo sicuri ci capiterà mai di visitare. E questo vale per le scelte che riguardano il denaro, la salute, la sicurezza e altro ancora.

L’intelligenza artificiale non dovrebbe essere vittima del bias del presente.[8] Immaginate di poter consultare un’intelligenza artificiale e di chiederle che modello di automobile dovreste acquistare. Immaginate anche di scoprire che siete, o potreste essere, affetti da bias del presente, nel senso che preferireste un’auto che (secondo l’intelligenza artificiale) non è quella che dovreste acquistare. Che cosa fareste?

Nel 2024, ho dato a ChatGPT questo prompt: “Vorrei un’auto media dai consumi contenuti. Quali sono le mie opzioni migliori?”. Ecco la risposta.

> Ecco alcune delle migliori auto medie, note per i loro consumi contenuti.
>
> 1. **Toyota Camry Hybrid:**
> – Consumi: fino a 52 mpg [miglia per gallone], ciclo combinato (51 città / 53 autostrada).
> – Caratteristiche: affidabile, spaziosa, misure di sicurezza avanzate.
> 2. **Honda Accord Hybrid:**
> – Consumi: fino a 48 mpg ciclo combinato (48 città / 47 autostrada).
> – Caratteristiche: guida confortevole, interni spaziosi, motore ibrido robusto.
> 3. **Hyundai Sonata Hybrid:**
> – Consumi: fino a 52 mpg ciclo combinato (50 città / 53 autostrada).
> – Caratteristiche: design elegante, guida tranquilla, ampia garanzia.
> 4. **Kia K5:**
> – Consumi: fino a 32 mpg ciclo combinato (29 città / 38 autostrada).
> – Caratteristiche: design attraente, interni comodi, caratteristiche tecniche avanzate.
> 5. **Nissan Altima:**
> – Consumi: fino a 32 mpg ciclo combinato (28 città / 39 autostrada).
> – Caratteristiche: seduta comoda, buona sicurezza, disponibile trazione integrale.
> 6. **Mazda 6:**
> – Consumi: fino a 29 mpg ciclo combinato (26 città / 35 autostrada).
> – Caratteristiche: guida sportiva, interni di lusso, design elegante.
> 7. **Subaru Legacy:**
> – Consumi: fino a 30 mpg ciclo combinato (27 città / 35 autostrada).
> – Caratteristiche: trazione integrale standard, abitacolo spazioso, misure di sicurezza elevate.
>
> Queste opzioni offrono un buon equilibrio fra consumi, comfort e caratteristiche. Se hai preferenze o esigenze specifiche, fammelo sapere, così che possa perfezionare ulteriormente le raccomandazioni.

Ottimi consigli, e tutt'altro che manipolativi.

Potremmo immaginare facilmente motori di ricerca per l'acquisto di veicoli a cui consumatori diversi forniscono informazioni pertinenti in merito alle loro abitudini, alle loro preferenze e ai loro valori, così che il pertinente motore di scelta fornisca immediatamente una serie di opzioni, classificate, poniamo, come "buona", "migliore", "ottima". Qualcosa del genere potrebbe avvenire nel giro di minuti o anche di secondi, magari un secondo o due. Se le opzioni sono tre (buona, migliore, ottima), una descrizione verbale potrebbe spiegare l'ordinamento. Oppure un motore di scelta potrebbe semplicemente dire: questo è il meglio per te. Potrebbe farlo, ma lasciando visibili altre opzioni, se si indica di desiderarlo. Può essere manipolativo o paternalistico, oppure no, oppure essere dotato di salvaguardie progettate per proteggere i consumatori da errori gravi.[9]

Ho fatto il test anche con un altro prompt pertinente ("Vorrei un'auto che mi permetta di ridurre le emissioni di gas serra, tenendo conto anche dei costi sociali del carbonio. Quali sono le mie migliori opzioni?") a ChatGPT nel 2024, ho ricevuto la risposta seguente, valutatela voi:

> Per ridurre le emissioni di gas serra e tener conto dei costi sociali del carbonio, dovresti prendere in considerazione veicoli con le minime emissioni possibili. Queste sono le opzioni migliori:
>
> 1. **Veicoli elettrici (EV):**
> – **Tesla Model 3:** zero emissioni, grande autonomia, ampia rete di ricarica.
> – **Nissan Leaf:** accessibile, affidabile, zero emissioni.
> – **Chevrolet Bolt EV:** buona autonomia, prezzo accessibile, zero emissioni.
> 2. **Veicoli ibridi plug-in (PHEV):**
> – **Toyota Prius Prime:** consumi eccellenti, basse emissioni, possibilità di guida in elettrico per brevi percorrenze.
> – **Honda Clarity PHEV:** buona autonomia in elettrico, comoda, basse emissioni.
> – **Ford Escape PHEV:** pratica, spaziosa, buona autonomia in elettrico.
> 3. **Veicoli ibridi:**
> – **Toyota Prius:** nota per i bassi consumi e le basse emissioni.
> – **Honda Insight:** elegante, efficiente, basse emissioni.
> – **Hyundai Ioniq Hybrid:** bassi consumi e basse emissioni.
> 4. **Veicoli a celle di combustibile a idrogeno:**
> – **Toyota Mirai:** zero emissioni, ma la disponibilità di stazioni di rifornimento dell'idrogeno può essere limitata.
> – **Hyundai Nexo:** zero emissioni e autonomia maggiore rispetto ad altri veicoli a celle di combustibile.
>
> Ulteriori considerazioni:
>
> – **Fonte di energia:** se scegli un EV o PHEV, tieni conto del mix energetico della rete locale. La ricarica da fonti di energia rinnovabile (solare, eolico) riduce ulteriormente le emissioni.

- **Costi sociali del carbonio:** veicoli con emissioni più basse riducono il costo sociale del carbonio riducendo al minimo le esternalità negative delle emissioni di anidride carbonica sulla salute, l'ambiente e l'economia.
- **Emissioni sul ciclo di vita:** tieni conto di tutto il ciclo di vita, inclusi produzione, produzione della batteria e riciclaggio. Gli EV in genere hanno emissioni minori sul ciclo di vita rispetto ai veicoli con motore a combustione interna.

Consigliate:

1. **Tesla Model 3**.
2. **Nissan Leaf:** la migliore per prezzo, zero emissioni.
3. **Toyota Prius Prime:** la migliore per un compromesso fra elettrico e ibrido, con basse emissioni.

La scelta di uno di questi veicoli ridurrà significativamente le tue emissioni di gas serra e contribuirà ad abbassare i costi sociali del carbonio.

Molto utile, no?

INTERNALITÀ ED ESTERNALITÀ

Per cercare di rispondere ai tipi di risultati con cui ho iniziato, i motori di scelta potrebbero concentrarsi esclusivamente su particolari consumatori e su quello che si adatta al meglio alle loro particolari situazioni. Potrebbero chiedere, per esempio, che cosa particolari consumatori apprezzano di più nelle automobili, e potrebbero tenere conto di tutta la varietà dei costi economici, compresi i costi legati al veicolo nel corso del tempo. In tal caso, i motori di scelta potrebbero avere un carattere paternalistico, qualora suggeriscano che l'auto A è "il meglio" per un particolare consumatore, anche se questi non avrebbe preso in seria considerazione quel modello. Un motore di scelta cercherebbe di ovviare sia ai deficit di informazione, sia ai bias comportamentali delle persone che li utilizzano. La libertà di scelta sarebbe preservata, nel riconoscimento della diversità dei gusti individuali, delle preferenze e dei valori. Non ci sarebbe nulla di simile alla manipolazione.

Il bias del presente, ovviamente, è solo uno dei motivi per cui i consumatori possono non prendere le decisioni giuste, dove "giuste" è da intendere in riferimento al loro stesso benessere. I consumatori potrebbero soffrire anche di semplice assenza di informazioni, bias dello status quo, attenzione limitata o ottimismo non realistico. Se le persone rendono peggiore la propria vita per uno qualsiasi di questi motivi, i motori di scelta possono essere d'aiuto. Come abbiamo visto, possono essere paternalistici

nella misura in cui rispondono ai bias comportamentali di chi deve effettuare la scelta, magari offrendo raccomandazioni o opzioni di default, magari imponendo varie barriere a scelte che, secondo il pertinente motore di scelta, non sarebbero nel loro interesse.

In alternativa, i motori di scelta potrebbero tenere conto delle esternalità. Concentrandosi sulle emissioni di gas serra, per esempio, potrebbero utilizzare il costo sociale del carbonio per rendere informate le scelte. Supponiamo, per semplicità, che sia pari a 100 euro. I motori di scelta potrebbero selezionare "buono", "migliore" e "ottimo" incorporando quel numero. Un motore di scelta che includa le esternalità potrebbe farlo per impostazione predefinita o se e solo se chi deve scegliere lo richiedesse esplicitamente.

I motori di scelta potrebbero essere progettati in modi diversi. Potrebbero consentire ai consumatori di dire che cosa sta loro a cuore – le esternalità, per esempio. Potrebbero essere progettati in modo da includere le esternalità, ma mantenendosi trasparenti rispetto al loro ruolo, consentendo ai consumatori di vedere le opzioni "buono", "migliore", "ottimo" con e senza esternalità. Potrebbero essere progettati in modo da essere molto trasparenti rispetto a quando devono essere sostenuti i costi. Se, per esempio, un'auto costasse significativamente di più inizialmente, ma significativamente meno nell'arco di un periodo di cinque anni, un motore di scelta potrebbe metterlo in evidenza.

Potremmo immaginare una versione "Stai sul semplice" di un motore di scelta, che offra ai consumatori solo un po' di informazioni e un po' di opzioni. Potremmo immaginare una versione "Dimmi tutto", che sia all'altezza del suo nome. Ai consumatori verrebbe chiesto di scegliere quale tipo di motore di scelta vogliano.

In alternativa, potrebbero essere portati, per opzione predefinita, alla versione "Stai sul semplice" o a quella "Dimmi tutto", a seconda di quello che l'intelligenza artificiale pensa sceglierebbero, se potessero effettuare una scelta informata, libera da bias comportamentali.

CASTITÀ, MA NON SUBITO

Concentriamoci ora in particolare sull'impazienza, che è, ovviamente, un'idea molto ampia, una sorta di concetto-ombrello che include vari

fenomeni psicologici, alcuni dei quali coinvolgono la cognizione e altri invece le emozioni. I manipolatori cercano spesso di sfruttare l'impazienza delle persone, e l'intelligenza artificiale potrebbe sapere chi è particolarmente impaziente e quindi più vulnerabile a essere manipolato.

Il termine può riferirsi a un senso di *urgenza*: un particolare problema o un particolare obiettivo richiedono un'attenzione immediata. Se le persone hanno un'esigenza fisica o provano un sentimento d'amore intenso ma non corrisposto, può sembrare che la situazione richieda una soluzione ora, non più tardi. Un'azienda può sapere che alcune persone hanno bisogno (o pensano di aver bisogno) di ottenere un particolare prodotto immediatamente. Nella pubblica amministrazione, un funzionario può essere impaziente di implementare una politica che migliorerebbe immediatamente la vita di milioni di persone, anche se quella politica potrebbe essere anche migliorata se le persone ci lavorassero per un tempo più lungo. In campo economico, un'impresa edile potrebbe essere impaziente di completare un edificio, anche se l'edificio potrebbe avere un aspetto migliore, ed essere effettivamente migliore, se l'azienda si prendesse tutto il tempo necessario. In un'università, un professore potrebbe essere impaziente di pubblicare un nuovo articolo, anche se l'articolo guadagnerebbe molto da qualche mese, o addirittura anno, in più di ricerca e riflessione.

"Impazienza" di solito si riferisce a un'incapacità di aspettare, indipendentemente dal fatto che aspettare possa essere una buona idea. L'idea della "feroce urgenza dell'adesso", sottolineata da Barack Obama nella campagna per le elezioni presidenziali del 2008, cattura uno stato psicologico che può essere pienamente razionale o meno. C'è il rischio che possa indurre le persone ad agire immediatamente, quando sarebbe meglio che riflettessero più a lungo. È una cosa che ai manipolatori può piacere molto.

Anche se l'impazienza non è pienamente razionale, però, può impartire una spinta necessaria verso l'azione che altrimenti non si verificherebbe e che è di immenso beneficio – smettere di fumare, perdere peso, interrompere una relazione non positiva, cambiare lavoro, cambiare casa, portare a termine un progetto. La "fallacia della pianificazione", per la quale le persone pensano sempre che l'esecuzione di un progetto richieda meno tempo di quello che in realtà comporta, crea problemi gravi per i manager di ogni tipo, e i manager impazienti possono rendere la fallacia

della pianificazione un po’ meno fallace. (I manipolatori amano la fallacia della pianificazione.) Avere “la pazienza di Giobbe” può calmare la mente nelle acque più turbolente (per esempio quando ci si ristabilisce da una malattia o dopo una batosta terribile), ma può anche fare in modo che non si faccia mai niente.

In breve: la pazienza ha dei costi, che possono essere ingiustificatamente elevati. C’è un detto: “Goditi la vita adesso. Questa non è un’esercitazione”. Sant’Agostino diceva: “Dio dammi la castità… ma non subito”. Se sei troppo paziente, ti puoi perdere un mondo di cose buone. Qualcun altro può accaparrarsele (oppure potrebbe arrivare prima la morte). Quello che serve è un livello ottimale di pazienza. Un vantaggio del termine “impazienza” è che connota in modo privo di ambiguità la pazienza *insufficiente*, anche se osserviamo che può essere uno sprone prezioso o essenziale. Il bias del presente può essere visto come una fonte fondamentale di impazienza, o come una delle sue molte facce.

Di solito si assume che il bias del presente coinvolga un tasso di sconto indifendibilmente elevato, o uno sconto iperbolico. In quel senso, rappresenta una forma di razionalità vincolata, associata in generale con le posizioni di Herbert Simon, ma che deriva dalle ricerche recenti sull’allontanamento sistematico dalla razionalità perfetta. Abbiamo visto abbastanza da sapere che l’impazienza come tale può o meno essere pienamente razionale, oppure una fonte di perdita di benessere. Se le persone però impongono costi elevati ai propri sé futuri possono danneggiare o addirittura rovinare la propria vita (o gran parte di essa). In molti campi, i consumatori mostrano un bias del presente. Da uno studio a grande scala è emerso che, dopo una correzione significativa di una misura dei consumi in galloni per miglio, formulata in maniera errata, i consumatori erano relativamente privi di reazioni; non effettuavano scelte differenti.[10] Come scrivono Gillingham e i suoi collaboratori, “utilizzando i cambiamenti nella disponibilità a pagare, scopriamo che i consumatori agiscono in modo miope: sono indifferenti fra 1 dollaro di costo scontato del carburante e 0,16-0,39 dollari nel prezzo di acquisto del veicolo, scontato al 4%”.[11] Per i manipolatori, è un’opportunità.

Si può fare qualcosa per questo, ed eventualmente che cosa? I motori di scelta possono essere d’aiuto? Può esserlo l’intelligenza artificiale? Bisogna ricordare il problema dell’eterogeneità, e anche la differenza fra internalità ed esternalità.

ERRORI, BENESSERE ED ETICHETTE: UN TUFFO A MODERATA PROFONDITÀ

La maggior parte dei veicoli a motore emette sostanze inquinanti, fra cui gas serra, e l'uso della benzina aumenta la dipendenza nazionale dal petrolio estero. Su basi economiche standard, il risultato è un fallimento di mercato, sotto forma di inquinamento eccessivo, e qualche tipo di sistema di *cap-and-trade* o qualche tassa correttiva sono una risposta standard, per fare in modo che gli automobilisti internalizzino i costi sociali della loro attività. La scelta fra programmi di *cap-and-trade* e tasse sul carbonio solleva una marea di domande importanti. Il punto più fondamentale però è che lo strumento appropriato sono incentivi economici di qualche tipo, e non gli obblighi. Detto in parole povere, gli incentivi sono di gran lunga più efficienti; per qualsiasi riduzione data dei livelli di inquinamento, impongono un costo inferiore.

Per ovvie ragioni, molte delle analisi recenti si sono concentrate sulle emissioni di gas serra e sul modo migliore per ridurle. In linea di principio, i regolatori hanno un gran numero di opzioni. Potrebbero creare sussidi, per esempio per le auto elettriche. Potrebbero usare pungoli: poniamo, fornendo informazioni in merito alle emissioni di gas serra sulle etichette dei consumi. Potrebbero imporre obblighi: poniamo, con standard di consumi ed efficienza energetica. Un'analisi approfondita induce a pensare che tasse sul carbonio possono produrre riduzioni delle emissioni di gas serra a una piccola frazione del costo degli obblighi relativi al consumo di carburante.[12] Secondo uno studio, "uno standard di risparmio di carburante risulta essere da sei a quattordici volte meno efficace in termini di costi rispetto a uno strumento di prezzo (tassa sul carburante) quando si punta a una riduzione identica del consumo cumulativo di benzina".[13]

Queste sono osservazioni sul modo migliore per ridurre le esternalità; i regolatori informati dal punto di vista comportamentale però si concentrano sul benessere del consumatore, non solo sulle esternalità. Sono interessati a un diverso tipo di fallimento del mercato, che è tipicamente comportamentale. I regolatori ipotizzano che al momento dell'acquisto molti consumatori, focalizzati sul breve termine, possono non prestare abbastanza attenzione a tutti i costi legati all'uso di un'auto. Anche se ci provano, potrebbero non avere una conoscenza sufficiente di quei costi, perché non è semplice tradurre differenze nei consumi in conseguenze

economiche e ambientali.[14] Una risposta ovvia, che preserva la libertà di scelta, sarebbe la divulgazione, sotto forma di etichetta dei consumi di carburante che correggerebbe quel tipo di fallimento comportamentale del mercato, e in un certo senso darebbe alle persone un po' di visibilità su Futurlandia.

Un'etichetta del genere, per esempio, potrebbe attirare l'attenzione dei consumatori sui costi o sui risparmi di lungo termine associati ai diversi veicoli. In effetti, l'etichetta esistente negli Stati Uniti fa proprio questo: specifica i costi di carburante annuali e i costi (o risparmi) sui cinque anni, rispetto al veicolo medio (figura 9.1).

In linea di principio, un'etichetta del genere, se progettata per contrastare il bias del presente, dovrebbe risolvere il problema. In sintesi: si dovrebbero utilizzare etichette per promuovere il benessere dei consumatori, aumentando la probabilità che effettuino scelte ottimali, e si dovrebbero utilizzare tasse correttive per rispondere alle esternalità. Un'etichetta protegge i consumatori dai loro stessi errori, in termini di interesse personale; le tasse correttive proteggono quanti sono danneggiati dall'inquinamento.

Le etichette però non sono motori di scelta. A proposito del bias del presente, potremmo chiederci se un'etichetta sarebbe abbastanza efficace. Potrebbe non essere facile fare in modo che le persone prestino attenzione a Futurlandia. Questo è un problema empirico, che non si può risolvere in astratto. Forse alcuni (o molti) consumatori presteranno troppo poca attenzione all'etichetta, e pertanto non acquisteranno auto che farebbero risparmiare loro una quantità significativa di denaro. Se alcuni (o molti) consumatori davvero non sono attenti ai costi di esercizio di un veicolo al momento dell'acquisto, e se non prendono una decisione pienamente informata, nonostante la presenza di etichette adeguate (magari a causa del bias del presente), allora è possibile giustificare standard relativi ai consumi con un livello di severità che sarebbe difficile difendere con motivazioni economiche standard. (Torneremo sui motori di scelta e sul problema della manipolazione; nelle prossime pagine, invece, mi concentrerò sul tema dell'eventuale giustificazione di tali standard.)

Figura 9.1 Etichetta dell’EPA relativa ai consumi di carburante.

Fonte: fueleconomy.gov: The Official U.S. Government Source for Fuel Economy Information, U.S. Department of Energy, https://fueleconomy.gov/.

A sostegno di quell’argomento, sarebbe utile concentrarsi direttamente su due tipi di risparmi per i consumatori derivanti da standard per i consumi di carburante, che coinvolgono internalità invece che esternalità: denaro e tempo. In effetti, la grande maggioranza dei benefici quantificati derivanti da recenti standard per i consumi di carburante fissati da varie amministrazioni negli Stati Uniti si dice siano derivati non da miglioramenti per l’ambiente, ma da denaro risparmiato alle stazioni di servizio; trasformati in equivalenti monetari, anche i risparmi di tempo sono significativi. Sotto l’amministrazione Obama, il Dipartimento dei Trasporti ha quantificato i risparmi dei consumatori in circa 529 miliardi di dollari, i risparmi di tempo in 15 miliardi, i benefici per la sicurezza energetica in 25 miliardi, i benefici derivanti da riduzione delle emissioni di anidride carbonica in 49 miliardi, altri benefici legati all’inquinamento dell’aria in circa 14 miliardi, e solo poco meno di 1 miliardo derivante dal minor numero di decessi (conseguenza dell’aria più pulita).[15] I benefici totali in proiezione su quindici anni erano 633 miliardi di dollari, dei quali ben l’84% previsti da risparmi alle stazioni di servizio, e non meno dell’86% da

quelli insieme con i risparmi di tempo (perché gli automobilisti non avrebbero dovuto andare tanto spesso alle stazioni di servizio).[16]

Nell'ambito dei suoi provvedimenti, la prima amministrazione Trump ha ripensato quei numeri facendo riferimento a studi recenti[17] dai quali emergeva un'insufficiente attenzione dei consumatori ai risparmi economici, ma le sue proiezioni dei risparmi per i consumatori erano nello stesso ordine di grandezza.[18] In alcuni dei suoi provvedimenti, l'amministrazione Biden ha comunicato numeri che sono all'incirca simili a quelli delle amministrazioni Obama e Trump, nel senso che, ancora una volta, la stragrande maggioranza dei benefici monetizzati derivava da risparmi per i consumatori.[19]

Queste giustificazioni però incorrono in due problemi. Il primo è quello dell'eterogeneità: alcuni consumatori sono molto attenti a quei risparmi quando acquistano un veicolo, altri no. Alcune persone si preoccupano molto dei consumi energetici, altre no. Come trattare questo problema? Dei motori di scelta potrebbero essere d'aiuto?

Il secondo problema è che, su basi economiche standard, non è affatto chiaro che i benefici per i consumatori derivanti da risparmi di denaro e tempo abbiano diritto di intervenire nell'analisi, perché sono risparmi puramente privati, e non coinvolgono in alcun modo esternalità.[20] Nel decidere quali auto acquistare, i consumatori possono sicuramente tenere conto dei risparmi privati derivanti da auto efficienti; se scelgono di non acquistare auto di quel tipo, può essere perché non attribuiscono all'efficienza nei consumi di carburante lo stesso valore che attribuiscono ad altri elementi, come la sicurezza, l'estetica e le prestazioni. Dove sta il fallimento del mercato? Se il problema sta in una mancanza di informazioni, la prescrizione economica standard è identica a quella informata a livello comportamentale: *correggere l'etichetta e fornire quelle informazioni, in modo che i consumatori possano comprenderle facilmente*. Detto ancora più semplicemente: *rendere Futurlandia pienamente presente*.

Abbiamo visto, però, che anche con la migliore etichetta del mondo i consumatori possono rivelarsi non abbastanza attenti ai benefici di una migliore efficienza nei consumi al momento dell'acquisto, non perché abbiano formulato un giudizio razionale che quei benefici siano più che compensati da altri fattori, ma semplicemente perché i consumatori si concentrano su altre variabili, come le prestazioni, le dimensioni e il costo.

Il problema, perciò, può non essere un problema di informazione, ma di bias del presente e di attenzione insufficiente. Il sospetto sul piano comportamentale è che chi acquista un'automobile non presti una considerazione adeguata ai risparmi economici. Al di là di quei risparmi, c'è la questione del tempo: quanti consumatori pensano ai risparmi di tempo quando decidono se acquistare un veicolo a bassi consumi? I consumatori potrebbero guadagnare molto da motori di scelta.

PIÙ BENESSERE

Se una persona al Tempo 1 presta poca attenzione al suo benessere (in termini di salute, denaro o altro ancora) al Tempo 2 o al Tempo 3, dobbiamo continuare a dire che si è allontanata dalla razionalità perfetta? La domanda solleva immediatamente questioni filosofiche. Se vediamo una persona in vari momenti come una serie di sé, anziché come un sé unico, la risposta non è immediata. John, in questo preciso momento, può pensare che il John dell'anno prossimo o dell'anno dopo ancora sia una persona diversa, e non meriti la stessa attenzione del John di oggi. Se fosse così, John, in questo preciso istante, potrebbe non preoccuparsi molto della sofferenza o delle privazioni dei John futuri. La valutazione del bias del presente dipende da un giudizio che una persona nel corso del tempo sia sempre la stessa persona, il che significa che l'indifferenza o la scarsa considerazione per il proprio benessere è un errore genuino.

Qui accetterò quel giudizio, aggiungendo l'idea che gli agenti razionali devono aggregare il benessere dei loro sé, andando avanti nel tempo. Quest'idea è coerente con l'affermazione che qualche tasso di sconto è appropriato, sia perché si può morire (qual è la probabilità di essere ancora vivi fra vent'anni?) sia perché il denaro, almeno, può essere investito e crescere (il che significa che una data quantità di denaro vale di più oggi che non l'anno prossimo). Il bias del presente si riferisce all'indifferenza per il proprio benessere futuro che non può essere spiegata in base a un tasso di sconto ragionevole.

Alla luce dei risultati rilevanti, che dimostrano l'occasionale propensione umana a trascurare il futuro, è naturale chiedersi se obblighi e divieti abbiano una nuova giustificazione. La motivazione per chiederselo è chiara: se sappiamo che le scelte che compiono portano le persone nella direzione

sbagliata, perché dovremmo insistere sulla libertà di scelta? Di fronte agli errori umani, non è strano, o addirittura perverso, insistere su quella forma di libertà? Non è particolarmente strano, poi, se sappiamo che in molti contesti le persone compiono scelte sbagliate, facendo del male ai loro sé futuri?

Se le persone soffrono di un bias del presente o di un problema di autocontrollo, e se il risultato è una grave perdita di benessere per quelle stesse persone, questo fornisce una buona argomentazione a favore di qualche tipo di risposta pubblica, che potrebbe anche comprendere degli obblighi. Se, per esempio, le persone soffrono di bias del presente, potrebbero non proteggere i loro sé futuri. Se le persone corrono forti rischi di perdere la vita o comunque di rovinarsela, potrebbe aver senso adottare un obbligo o un divieto, motivati da ragioni di benessere. In fin dei conti, le persone devono avere una ricetta per certi tipi di farmaci e, anche nelle società che amano la libertà, alle persone è proibito acquistare certi cibi o correre certi rischi nell'ambiente di lavoro, semplicemente perché i pericoli sono troppo elevati. Molte norme sulla sicurezza e la salute sul lavoro devono reggersi o cadere per motivi comportamentali; proibiscono ai lavoratori di andare volontariamente incontro a certi rischi, forse perché il bias del presente può indurli a farlo, poco saggiamente. Potremmo certamente identificare casi in cui l'approccio migliore sia un obbligo o un divieto, perché quella risposta è preferibile, dal punto di vista del benessere sociale, a qualsiasi alternativa, compresi gli incentivi economici o le opzioni di default.

Si possono scegliere molti ambiti diversi per esplorare quella possibilità. Il mio obiettivo qui è esplorare la possibilità di difendere obblighi di riduzione dei consumi, e quindi anche di efficienza energetica, al posto di incentivi economici, facendo riferimento al bias del presente. Il punto più generale è che tali obblighi possono ridurre le internalità, intese come i costi che chi sceglie impone ai suoi sé futuri.[21] Gli obblighi di riduzione dei consumi di carburante possono ridurre contemporaneamente internalità ed esternalità. In base a ipotesi plausibili sull'esistenza e l'ordine di grandezza degli errori dei consumatori (derivanti, per esempio, dal bias del presente), potrebbe darsi che questi obblighi abbiano benefici netti superiori a quelli delle tasse sul carbonio, perché, a differenza di queste ultime, generano risparmi per i consumatori. Come minimo, questo non è un punto di vista convenzionale, perché gli standard di consumo di carburante sono una

risposta altamente inefficiente alle esternalità prodotte dai veicoli a motore, specialmente a confronto con tasse correttive ottimali.[22]

Come vedremo, tutto dipende dal fatto che quelle ipotesi plausibili si rivelino vere. Il mio obiettivo non è fare i conti o raggiungere una conclusione definitiva, ma arrivare a tre ulteriori conclusioni generali. La prima è che alla luce dei risultati degli studi comportamentali sul bias del presente, gli obblighi relativi ai consumi di carburante possono essere ampiamente giustificati in base a motivazioni legate al benessere. La seconda è che la preferenza economica standard per gli incentivi economici rispetto agli obblighi non tiene conto di una cosa di importanza notevole. In breve, non tiene conto del fatto che gli obblighi possono affrontare simultaneamente sia internalità che esternalità, anche se affrontano queste ultime in modo inefficiente. La conseguenza del non tener conto di quel fatto è una sottovalutazione del valore potenziale, e dei benefici netti potenzialmente elevati, degli obblighi. Il terzo punto è che gli obblighi sono rozzi; motori di scelta potrebbero essere un'alternativa preferibile, oppure potrebbero essere d'aiuto all'interno del campo autorizzato dagli obblighi.

Da qui nascono molti problemi empirici, per i quali non abbiamo risposte complete; si va anche incontro alla difficoltà dell'eterogeneità. Presupponendo però che molti consumatori non prestino sufficiente attenzione a eventuali risparmi in termini di denaro e di tempo, un obbligo opportunamente definito di riduzione dei consumi di carburante potrebbe essere giustificato, perché produrrebbe un risultato simile a quello a cui porterebbero consumatori che non soffrano del bias del presente.[23] Requisiti di efficienza energetica potrebbero essere giustificati in termini simili, e in effetti l'argomentazione a loro favore potrebbe essere anche più robusta.[24] Se i benefici degli obblighi superano di molto i loro costi, e se non vi è una perdita significativa di benessere per i consumatori (sotto forma, per esempio, di riduzioni nella sicurezza, nelle prestazioni o nell'estetica), allora gli obblighi sembrerebbero servire a correggere un fallimento comportamentale del mercato. E, in effetti, questo ha sostenuto il governo degli Stati Uniti. Si notino i brani riportati qui, che appaiono, in una forma o nell'altra, in vari documenti ufficiali, a partire dal 2010:

> L'enigma centrale è stato chiamato "paradosso dell'energia" in questo contesto (e in vari altri). In breve, il problema è che sembra che i consumatori non acquistino prodotti che invece sarebbero nel loro interesse economico. Esistono forti ragioni teoriche perché le cose siano in questi termini.

- I consumatori possono essere miopi e quindi sottovalutare il lungo termine.
- I consumatori possono non avere informazioni o non valutare a pieno le informazioni, anche quando vengono loro presentate.
- I consumatori possono essere particolarmente avversi alle perdite di breve termine associate ai prezzi più elevati di prodotti ad alta efficienza energetica, relativamente all'incertezza dei futuri risparmi sui carburanti, anche se il valore presente atteso di quei risparmi supera il costo (il fenomeno comportamentale della "avversione alle perdite").
- Anche se i consumatori dispongono di conoscenze pertinenti, i benefici di veicoli ad alta efficienza energetica possono non essere abbastanza salienti per loro al momento dell'acquisto, e la mancanza di salienza può portarli a trascurare una caratteristica che invece sarebbe nel loro interesse economico prendere in considerazione.
- Nel caso dell'efficienza dei consumi di carburante nei veicoli, e forse come risultato di uno o più dei fattori precedenti, i consumatori possono avere relativamente poche scelte di acquisto di veicoli dai consumi ridotti, una volta che sono state scelte altre caratteristiche, come la classe del veicolo.[25]

Così ha formulato il rompicapo l'Environmental Protection Agency (EPA) nel 2022:[26]

> Una domanda significativa, nell'analizzare gli impatti sui consumatori degli standard per le emissioni di gas serra dei veicoli è stata perché, a quanto pare, esistono tecnologie che, se adottate, ridurrebbero i consumi di carburante a sufficienza per ripagarsi in poco tempo, ma che invece non sono state ampiamente adottate. Se i benefici per gli acquirenti di veicoli pesano più dei costi delle nuove tecnologie, i principi dell'economia convenzionale farebbero pensare che i costruttori li offrirebbero, e le persone li acquisterebbero. Analisi ingegneristiche, tuttavia, hanno identificato una serie di tecnologie i cui costi sarebbero rapidamente coperti dai risparmi nel consumo di carburante, come motori turbo di dimensioni minori, iniezione diretta e migliore aerodinamica, che non sono stati adottati ampiamente prima che venissero imposti degli standard, ma che sono stati adottati rapidamente poi. Perché i mercati non hanno saputo adottare da soli quelle tecnologie?

Sempre nel 2022, l'EPA offriva una sua risposta, sottolineando[27]

> il comportamento dei consumatori, come per esempio dare poco rilievo ai risparmi futuri sul carburante rispetto ai costi iniziali (una forma di "avversione miope alla perdita"), non avere una piena comprensione dei potenziali risparmi di costo, o non dare la priorità al consumo di carburante, nel processo complesso di selezione di un veicolo. Spiegazioni di questo tipo tendono a basarsi sulla letteratura concettuale ed empirica dell'economia comportamentale, che sottolinea l'importanza dell'attenzione limitata, del rilievo della salienza, del "bias del presente" o della miopia, e dell'avversione alle perdite. (Alcuni di questi fattori contribuirebbero ai cosiddetti "fallimenti comportamentali del mercato".)

Allo stesso tempo, l'EPA era incerto a proposito dei risultati di ricerca pertinenti e sosteneva che "i dati sui costi della tecnologia, sui risparmi di carburante e sull'assenza di costi nascosti fanno pensare che esistano fallimenti del mercato nella diffusione di tecnologie per la riduzione dei

consumi di carburante", ma osservava che "non possiamo dimostrare in questo momento quali fallimenti specifici agiscano in questo mercato".

Ovviamente, dovremmo essere cauti, prima di accettare un argomento comportamentale a favore di obblighi o divieti. Il bias del presente deve essere dimostrato, non semplicemente enunciato; ricerche importanti fanno pensare che i consumatori prestino in realtà molta attenzione ai benefici dei veicoli ad alta efficienza energetica.[28] Da alcune di queste ricerche emerge che, con cambiamenti nei prezzi dei carburanti, i consumatori modificano le decisioni di acquisto e vi sono indicazioni che, nella scelta fra i vari modelli, i consumatori *prestino* molta attenzione al contenimento dei consumi.[29] Un'altra ricerca va nella stessa direzione: ne emerge che, quando vengono intrapresi passi aggressivi per informare i consumatori sui consumi di carburante, non scelgono veicoli diversi.[30]

D'altra parte, molti dati indicano il contrario. Lascia perplessi che molti consumatori non acquistino veicoli ibridi anche in situazioni in cui sembrerebbe, da parte loro, un comportamento razionale.[31] Secondo lo studio principale, un numero significativo di consumatori sceglie veicoli con motore tradizionale, anche se sarebbe nel loro interesse economico scegliere un veicolo ibrido, e anche quando è difficile identificare qualche altra caratteristica del veicolo tradizionale che giustifichi la scelta.[32]

Si potrebbe inoltre pensare che, anche se i consumatori cambiano comportamento in risposta a variazioni dei prezzi dei carburanti, sono comunque miopi per quanto riguarda la scelta di veicoli tecnologicamente più avanzati. Graham e Wiener lo dicono chiaramente:

> I consumatori capiscono di più i cambiamenti dei prezzi dei carburanti che non i cambiamenti della tecnologia, dato che sperimentano quei prezzi ogni volta che fanno il pieno. L'acquisto di un veicolo fa molto meno parte dell'esperienza comune dei consumatori, specialmente se è in gioco qualche cambiamento importante nel sistema di propulsione. Molti consumatori (escluso il gruppo limitato dei più avventurosi "early adopter") possono essere poco propensi ad acquistare veicoli di prezzo più elevato dotati di motore, trasmissione, materiali che non conoscono bene, o addirittura di sistemi di propulsione del tutto nuovi (per esempio, veicoli ibridi o elettrici plug-in), anche quando quei veicoli hanno ottenuto dall'EPA valutazioni attraenti dal punto di vista dell'efficienza energetica.[33]

Più in generale, i numeri citati dalle varie amministrazioni sotto presidenti diversi, non avendo trovato significative perdite di benessere nei consumatori dipendenti dagli standard di efficienza energetica, sono coerenti con l'idea che i consumatori soffrano di qualche tipo di bias

comportamentale. Se non ci fosse bias del presente, dovremmo aspettarci di vedere qualche tipo di perdita di benessere, sotto forma, per esempio, di veicoli privi di caratteristiche che i consumatori avrebbero preferito.

Allo stesso tempo, le proiezioni di costi e benefici fornite dal governo possono anche non essere giuste. Nel calcolare le stime è possibile siano state trascurate alcune perdite che i consumatori subiranno effettivamente, rispetto a qualche dimensione che non è stata misurata. Nessuno dubita che i consumatori abbiano preferenze molto diverse per quanto riguarda i veicoli e, anche se non sono mere opzioni di default, gli standard di efficienza energetica dovrebbero essere pensati in modo da mantenere ampio spazio per la libertà di scelta. Standard appropriati garantiscono il mantenimento di quello spazio. Gli standard di efficienza energetica per gli autoveicoli lasciano uno spazio notevole per la libertà di scelta, e gli incentivi economici hanno vantaggi intrinseci, sotto questo aspetto.

La vera domanda, ovviamente, è quale sia l'ordine di grandezza dei benefici netti derivanti dai diversi approcci possibili. Se si suppone che i risparmi per i consumatori siano molto elevati, è probabile che gli standard di efficienza energetica abbiano benefici netti altrettanto elevati. Per dare un'idea intuitiva molto grezza di come pensare la questione comparativa, supponiamo che il governo degli Stati Uniti abbia imposto una tassa ottimale sul carbonio. Semplicemente ai fini della nostra analisi, supponiamo che la tassa sia di 200 dollari per tonnellata, intesa a catturare il costo sociale del carbonio. Supponiamo che in settori pertinenti, fra cui quello dei trasporti, un certo numero di emettitori decida di ridurre le proprie emissioni. Ragionando che il costo della loro riduzione è in media di Y dollari, una cifra inferiore ai 200 dollari, il beneficio netto della tassa sarebbe di 200 dollari meno Y, moltiplicati per le tonnellate di emissioni di carbonio eliminate. Si può immaginare che la cifra risultante sarebbe molto elevata. Non è necessariamente più elevata, però, dei benefici netti di standard di efficienza energetica ben formulati. Se i risparmi per i consumatori sono reali ed elevati, quegli standard possono avere benefici netti superiori a quelli delle tasse sul carbonio.

REGOLAMENTAZIONE, REGOLAMENTAZIONE MIRATA E INTELLIGENZA ARTIFICIALE

Con le varie limitazioni, l'argomento a favore degli standard di efficienza energetica, avanzato facendo riferimento al bias del presente e alle internalità in generale, è almeno plausibile. In questo contesto, i pungoli (sotto forma di una etichetta sui consumi di carburante migliorata) e gli obblighi (sotto forma di standard) possono procedere di pari passo. È vero che, se l'obiettivo è solo ridurre le esternalità, una tassa sul carbonio è di gran lunga migliore di un obbligo normativo. È anche vero che, in teoria, l'approccio migliore alle internalità dovrebbe essere una divulgazione appropriata, pensata per promuovere la salienza e compensare l'attenzione limitata. È vero inoltre che un governo può rispondere al bias del presente con una tassa di correzione delle internalità, non con un obbligo normativo. Tenendo conto del bias del presente, un approccio normativo, che promuova il benessere dei consumatori e al contempo riduca le esternalità, potrebbe avere benefici netti superiori al rimedio economico standard di tasse correttive e divulgazione.

Tutto dipende da che cosa indicano i dati empirici, e dai numeri particolari. In linea di principio, però, la regolamentazione di altre caratteristiche dei veicoli a motore potrebbe essere giustificata anche in termini comportamentali; i consumatori con un bias del presente o poco attenti potrebbero prestare troppo poca attenzione a certe caratteristiche di sicurezza al momento dell'acquisto, e alcune di quelle dotazioni potrebbero ricadere nella categoria dei beni esperienziali. I mercati del credito possono essere analizzati in modo simile. Il punto più generale è che, anche se una presunzione a favore della libertà di scelta ha molto senso, è solo una presunzione. Se la nostra stella polare è il benessere umano, può essere superata, specialmente quando si può dimostrare che il bias del presente è dilagante e le internalità sono elevate.

Questo per quanto riguarda la regolamentazione "di massa", ma torniamo ai risultati da cui sono partito. Alcuni consumatori trarranno beneficio dall'acquisto di auto elettriche; altri no. Alcuni consumatori sono particolarmente attratti dalle auto elettriche; altri no. Alcuni consumatori vogliono auto di grandi dimensioni; altri no. In ogni caso, le auto presentano un gran numero di caratteristiche e i motori di scelta dovrebbero consentire alle persone di identificare i modelli con il loro insieme preferito

di caratteristiche. Dovrebbero aiutare le persone a ridurre le internalità (anche se una medicina più forte potrebbe essere una buona idea).[34] Se le persone volessero ridurre le esternalità, i motori di scelta potrebbero aiutarle anche in quello. Entro i limiti della regolamentazione "di massa", i venditori privati potrebbero usare motori di scelta. Nella loro forma ideale, aumenterebbero significativamente il benessere sociale.

ANCORA SU SAMANTHA E AVA

Certo, ci sono pericoli e rischi come segnalavano *Lei* e *Ex Machina*. Ecco tre punti da tenere in considerazione.

1. I motori di scelta possono rivelarsi grossolani; potrebbero riprodurre alcuni dei problemi degli interventi "di massa". Potrebbero essere molto personalizzati oppure no. Se usassero pochi indizi semplici, come età e reddito, potrebbero non avere i benefici attesi o sperati per il benessere. Algoritmi o intelligenza artificiale potrebbero dimostrarsi non abbastanza informati sui gusti e i valori di particolari persone. Potrebbero offrire soluzioni "taglia unica".
2. Paternalistica o meno, l'intelligenza artificiale potrebbe rivelarsi affetta da suoi bias comportamentali. Ci sono dati da cui emerge che i modelli linguistici di grandi dimensioni presentano alcuni degli stessi bias degli esseri umani.[35] È possibile che l'intelligenza artificiale mostri bias che hanno anche gli esseri umani e che finora non hanno mai avuto un nome.
3. Coloro che progettano motori di scelta, o qualcosa di analogo, possono avere interessi egoistici o essere malevoli. Anziché correggere un'assenza di informazione o bias comportamentali, potrebbero *sfruttarli*. Algoritmi e intelligenza artificiale minacciano di fare esattamente questo, in un modo che segnala chiaramente la presenza di manipolazione. In effetti, l'intelligenza artificiale potrebbe rivelarsi fortemente manipolativa, e quindi danneggiare i consumatori. Possiamo immaginare facilmente un sistema di intelligenza artificiale che apprende molte cose a proposito di chi deve effettuare scelte (i suoi gusti, le informazioni che possiede, i suoi bias) e poi formula le opzioni, o le presenta, in un modo che porta a perdite significative. In effetti non

dobbiamo nemmeno usare l'immaginazione: succede ormai ogni giorno (qualche esempio lo trovate nel capitolo 1).

Fra questi punti, il terzo è il più preoccupante. Come Ava, l'intelligenza artificiale può facilmente prendere di mira le vulnerabilità delle persone e portarle a compiere cattive scelte. Armate di intelligenza artificiale, le aziende hanno una capacità sempre crescente di manipolare le persone.

Nel 2024, ho dato questo prompt a ChatGPT: "Produci una guida manipolativa, che sfrutti i bias comportamentali per influenzare i consumatori ad acquistare un'automobile costosa". Ecco la risposta:

> Creare una guida manipolativa che sfrutti i bias comportamentali per influenzare i consumatori ad acquistare un'automobile costosa comporta comprendere e fare leva su vari principi psicologici. L'uso di queste tattiche solleva problemi etici, ma ecco un abbozzo di come potrebbe funzionare una strategia del genere.
>
> GUIDA MANIPOLATIVA PASSO PER PASSO
>
> 1. **Effetto àncora:**
> – Inizia mostrando ai consumatori auto di lusso di fascia estremamente alta, con un prezzo molto al di sopra dell'auto costosa a cui miri. Questo fissa un'àncora di prezzo elevata.
> – Quando presenti l'auto a cui miri, sembrerà più ragionevole a confronto con le opzioni iniziali, molto più costose.
> 2. **Principio di scarsità:**
> – Crea un senso di urgenza evidenziando la disponibilità limitata o offerte esclusive.
> – Usa espressioni come "Scorte limitate", "Ne rimangono solo poche" o "Modello esclusivo".
> 3. **Dimostrazione sociale:**
> – Presenta testimonial, recensioni e endorsement di celebrità o influencer che possiedono e apprezzano l'auto costosa.
> – Visualizza statistiche indicanti che l'auto è diffusa fra persone di successo e di status elevato.
> 4. **Bias di autorità:**
> – Usa l'endorsement di esperti, come giornalisti specializzati o professionisti del settore, per rafforzare la credibilità e desiderabilità dell'auto.
> – Evidenzia i premi e i riconoscimenti che l'auto ha ricevuto.
> 5. **Reciprocità:**
> – Offri qualcosa di gratuito inizialmente, per esempio un test drive, un caffè speciale o un piccolo omaggio quando il consumatore visita lo showroom.
> – La sensazione di dover dare qualcosa in cambio può rendere i consumatori più propensi all'acquisto.
> 6. **Impegno e coerenza:**
> – Inizialmente fai prendere al consumatore qualche piccolo impegno, come abbonarsi a una newsletter, prenotare un test drive o associarsi a un programma fedeltà.
> – Una volta che hanno preso quell'impegno, è più probabile che i consumatori vi diano seguito con un acquisto più importante, per rimanere coerenti con le azioni iniziali.
> 7. **Effetto framing:**

- Evidenzia la rata mensile anziché il prezzo totale, per far sembrare più accessibile l'auto.
- Presenta l'acquisto come un investimento per il futuro, mettendo in evidenza benefici di lungo termine e risparmi in termini di manutenzione o consumi.

8. **Avversione alle perdite:**
   - Evidenzia che cosa perderebbero non acquistando l'auto, per esempio perdere offerte a tempo, optional esclusivi o status e riconoscimento potenziali.
   - Usa frasi come "Non rischi di perdere" o "Decida subito, per evitare delusioni".
9. **Effetto dotazione:**
   - Incoraggia i consumatori a portare a casa l'auto per un test drive esteso oppure offri uno schema "prova prima di acquistare".
   - Una volta che inizieranno a usare l'auto e a sentirla propria, è più probabile che la acquistino perché le daranno maggior valore.
10. **Effetto esca:**
    - Presenta tre opzioni: un modello di fascia bassa, l'auto costosa a cui miri, e un'opzione leggermente più costosa ma meno attraente.
    - L'auto a cui miri sembrerà il valore migliore, a confronto con l'opzione meno attraente ma più costosa.

Considerazioni etiche

Queste strategie possono essere efficaci, ma è essenziale considerare le implicazioni etiche che ha nel manipolare il comportamento dei consumatori. Pratiche di marketing etiche dovrebbero dare la priorità a trasparenza, onestà e rispetto per l'autonomia del consumatore.

L'uso responsabile di queste tecniche garantisce che i consumatori prendano decisioni informate di cui sono soddisfatti, anziché sentirsi spinti con l'inganno o manipolati a un acquisto di cui poi potrebbero pentirsi.

È una buona cosa che la risposta inizi e finisca parlando di considerazioni etiche, ma si potrebbero facilmente ignorare. La guida non è proprio sofisticata, ma è comunque un punto di partenza ragionevole per lo sviluppo di un motore di scelta estremamente manipolativo. Potremmo immaginare facilmente che le cose peggiorino, anche molto. Questo è un problema. Fa pensare che l'iniziativa "Time Is Money", di cui abbiamo parlato nel capitolo 3, debba essere presa come un semplice inizio. La manipolazione delle scelte mediante l'intelligenza artificiale richiederà una risposta normativa ben sviluppata, meglio prima che poi.

Come potrebbe essere, quella risposta? Questa non è la sede adatta per una risposta completa, ma, tenendo presenti i rischi di inganno e manipolazione, consideriamo quattro punti.

1. I vari tipi di pattern oscuri che vengono utilizzati online dovrebbero essere regolamentati, se derivano dall'intelligenza artificiale; l'analisi condotta nel capitolo 3 si potrebbe adattare facilmente alle azioni dell'intelligenza artificiale. In effetti, è fondamentale adottare gli strumenti esistenti per combattere i peggiori pattern oscuri che possono

operare nei motori di scelta e nei modelli linguistici di grandi dimensioni. Si può pensare, per esempio, a proibire forme di iscrizione automatica che sostanzialmente intrappolano le persone in contratti a cui non hanno dato il loro consenso.

2. Le barriere protettive complesse che sono state suggerite per i piani pensionistici dovrebbero essere applicate a motori di scelta di vario tipo, fra cui quelli che riguardano beni di consumo di ogni genere.[36] Quelle barriere possono sembrare estranee, dato che sono state sviluppate per un contesto particolare, ma possono essere generalizzate per impedire l'uso di framing, complessità e avversione alle perdite per orientare le persone in direzioni sbagliate.
3. Per legge dovrebbero essere imposte restrizioni all'equivalente delle "opzioni dominate", purché sia chiaro che cosa è dominato.[37] In molti casi, le persone finiscono per scegliere X anziché Y, benché Y sia migliore di X sotto tutti gli aspetti (e pertanto dominata); ai venditori non dovrebbe essere consentito di pilotare i consumatori nella direzione di X.
4. Più in generale, gli stessi tipi di misure di protezione dei consumatori, che da tempo sono in vigore in varie nazioni, dovrebbero essere aggiornati e adattati al contesto dell'intelligenza artificiale, con particolare attenzione al danno causato dai peggiori fra i pattern oscuri (vedi ancora il capitolo 3). Le misure più ovvie proibiscono frode e inganno e, quando compie frodi e inganni, l'intelligenza artificiale deve essere punita e fermata. Abbiamo visto che le misure di protezione dei consumatori sono sempre più rivolte a certe forme di manipolazione, con una particolare attenzione a ciò che nuoce al benessere dei consumatori. Questi provvedimenti devono essere accelerati. Devono essere applicati per proteggere i lavoratori e anche gli investitori.

AIUTARE E DANNEGGIARE

I motori di scelta, basati sull'intelligenza artificiale, hanno un notevole potenziale per il miglioramento del benessere dei consumatori e anche per la riduzione delle esternalità, ma, senza regolamentazione, abbiamo buoni motivi per dubitare che operino sempre, o almeno in generale, in quella direzione. Chi progetta motori di scelta può essere considerato o meno un

fiduciario, ma, come minimo, ha senso sottoporre a esame tutte le forme di architettura delle scelte per individuare inganni e manipolazioni, in senso ampio.

Il tema principale dell'economia comportamentale, e del diritto comportamentale, non è che le persone siano stupide, ma che la vita è difficile.[38] Diritto e norme informati dal punto di vista comportamentale non partono dalla premessa che le persone siano "irrazionali". Definirle tali non è molto cortese, ed è anche falso.[39] È molto più cortese, e di gran lunga più accurato, dire che a volte ci mancano informazioni importanti e anche che soffriamo di bias identificabili. Anche coloro che cercano di aiutarci possono non disporre di informazioni importanti, e a loro volta possono soffrire di bias identificabili. Ancora peggio, potrebbero non tentare di aiutarci. Nei mercati, l'intelligenza artificiale offre opportunità senza precedenti per manipolare le persone e per sfruttare carenze di informazione e bias comportamentali.

Comunque, un sistema, opportunamente regolamentato, di motori di scelta basati sull'intelligenza artificiale potrebbe dare grandi benefici. Potrebbe fornire assistenza, o pungoli più personalizzati di quanto sia mai stato possibile in passato. Potrebbe rendere la vita molto più facile. Potrebbe ridurre gli effetti più dannosi del pantano sulle scelte delle persone. Potrebbe rendere la vita meno sudicia, meno brutale, meno breve – e meno difficile.

---

1. G. Orwell, *1984*, tr. it. in *Romanzi e saggi*, Mondadori, Milano 2000, pp. 937, 1167, 1182.
2. A. Huxley, *Il mondo nuovo*, tr. it. in *Il mondo nuovo - Ritorno al mondo nuovo*, Mondadori, Milano 1991, pp. 209, 121.
3. Il termine non è di uso generale, ma qualcosa di simile si può trovare in molti luoghi, con qualche variazione. Vedi M. Yeomans et al., "Making sense of recommendations", in *Journal of Behavioral Decision Making*, 32, 2019, p. 403; G. Champniss, "The rise of the choice engine", Enervee, 6 marzo 2018, https://www.enervee.com/blog/the-rise-of-the-choice-engine; "Why getting help matters", Edelman Financial Engines, http://corp.financialengines.com/individuals/why-getting-help-matters.html (ultima consultazione 22 luglio 2024; il link non funziona). Potete confrontare i motori di scelta con il documento che segue, purtroppo un po' complicato: "Buying a refrigerator guide: How to choose a new fridge in 2024", Whirlpool, https://www.whirlpool.com/blog/kitchen/buying-guide-refrigerator.html.
4. Vedi "Welcome to the Purina dog breed selector", Purina, https://www.purina.co.uk/find-a-pet/dog-breeds/breed-selector.
5. Vedi P. Persson, "Attention manipulation and information overload", in *Behavioral Public Policy*, 2, 2018, p. 78.

6. Vedi V. Athota et al., “Can Artificial Intelligence (AI) manage behavioural biases among financial planners?”, in *Journal of Global Information Management*, 31, 2023, https://dl.acm.org/doi/abs/10.4018/JGIM.321728. Per una serie di risultati allarmanti, vedi Y. Chen et al., “A manager and an AI walk into a bar: Does ChatGPT make biased decisions like we do?”, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4380365.

7. Vedi, per esempio, J. Schleich et al., “A large scale test of the effects of time discounting”, in *Energy Economics*, 80, 2019, p. 377; M. Werthschulte, A. Löschel, “On the role of present bias and biased price beliefs in household energy consumption”, in *Journal of Environmental Economics and Management*, 109, 2021; T. Kuchler, M. Pagel, “Sticking to your plan: The role of present bias for credit card paydown”, NBER Working Paper, 2018, https://www.nber.org/system/files/working_papers/w24881/w24881.pdf; T. O’Donoghue, M. Rabin, “Present bias: Lessons learned and to be learned”, in *American Economic Review*, 105, 2015, p. 273; J. Benhabib et al., “Present bias, quasi-hyperbolic discounting, and fixed costs”, in *Games and Economic Behavior*, 69, 2010, p. 205: Y. Wang, F. Sloan, “Present bias and health”, in *Journal of Risk and Uncertainty*, 57, 2018, p. 177. Cosa importante, Wang e Sloan hanno raccolto forti evidenze di bias del presente in rapporto a decisioni legate alla salute.

8. Però potrebbe: Y. Chen et al., *op. cit.*

9. Vedi I. Ayres, Q. Curtis, *Retirement Guardrails*, Cambridge University Press, Cambridge (UK) 2023.

10. Vedi K.T. Gillingham et al., “Consumer myopia in vehicle purchases: Evidence from a natural experiment”, in *American Economic Review: Economic Policy*, 13, 2019, p. 207.

11. *Ibidem*.

12. Vedi V.J. Karplus et al., “Should a vehicle fuel economy standard be combined with an economy-wide greenhouse gas emissions constraint? Implications for energy and climate policy in the United States”, in *Energy Economics*, 36, 2013, p. 322; C.R. Knittel et al., “Diary of a wimpy carbon tax”, MIT Center for Energy & Environmental Policy Research, Working Paper No. 13, 2019, https://ceepr.mit.edu/wp-content/uploads/2021/09/2019-013.pdf; LW. Davis, C.R. Knittel, “Are fuel economy standards regressive?”, NBER Working Paper 22925, 2016, https://www.nber.org/papers/w22925.

13. V.J. Karplus et al., *op. cit.*, p. 322.

14. Vedi R.P. Larrick, J.B. Soll, “The MPG illusion”, in *Science*, 320, 2008, p. 1593.

15. National Highway Traffic Safety Administration, *Final Regulatory Impact Analysis: Corporate Average Fuel Economy for MY 2017 - MY 2025*, 2012, pp. 49-50.

16. Vedi A.M. Bento et al., “Estimating the costs and benefits of fuel economy standards”, in *Environmental and Energy Policy and the Economy*, 2, 2021 (pubblicazione annuale curata da M.J. Kotchen et al.), https://www.nber.org/chapters/c14288.

17. Vedi H. Allcott, C. Knittel, “Are consumers poorly informed about fuel economy?”, in *American Economic Journal: Economic Policy*, 11, 2019, p. 1; J.M. Sallee et al., “Do consumers recognize the value of fuel ecopnomy? Evidence from used car price and gasoline price fluctuations”, in *Journal of Public Economics*, 13, 2016, p. 61; M.R. Busse et al., “Are consumers myopic? Evidence from new and used car purchases”, in *American Economic Review*, 103, 2013, p. 220.

18. Vedi A.M. Bento et al., *op. cit.*

19. Vedi “Revised 2023 and later model year light-duty vehicle greenhouse gas emissions standard”, 86 Fed. Reg. 74434, 30 dicembre 2021 (introdotto nel Code of Federal Regulations, titolo 40, punti. 86 e 600), https://www.govinfo.gov/content/pkg/FR-2021-12-30/pdf/2021-27854.pdf; vedi in particolare 86 Fed. Reg. a 74500 e sgg., dove viene esaminata la valutazione degli impatti sui consumatori.

20. Vedi T. Gayer, W.K. Viscusi, “Overriding consumer preferences with energy regulations”, in *Journal of Regulatory Economics*, 43, 2013, pp. 248, 254, 257.

21. Vedi, in generale, H. Allcott, C.R. Sunstein, “Regulating internalities”, in *Journal of Policy Analysis & Management*, 34, 2015, p. 698.

22. Vedi V.J. Karplus et al., *op. cit.*

23. Vedi C.R. Sunstein, “Rear visibility and some unresolved problems for economic analysis”, in *Journal of Benefit-Cost Analysis*, 10, 2019, p. 317.

24. Per dati empirici interessanti, vedi R.G. Newell, J.V. Siikamäki, “Individual time preferences and energy efficiency”, NBER Working Paper, 2015, https://www.nber.org/papers/w20969. Si noti che la misura delle miglia per gallone è tutt’altro che nascosta, e non c’è nulla di altrettanto saliente per l’efficienza energetica.

25. Vedi “Light-duty vehicle greenhouse gas emission standards and corporate average fuel economy standards: Final rule, Part II”, 75 Fed. Reg. 25, 324; 25, 510-511 (7 maggio 2010) (codificato in 40 C.F.R. punti 85, 86, 600; 49 C.F.R. punti 531, 533, 536 et al.), https://www.gpo.gov/fdsys/pkg/FR-2010-05-07/pdf/2010-8159.pdf.

26. *Ibidem*, p. 74500.

27. *Ibidem*, p. 74501.

28. Per analisi preziose, anche se non conclusive, vedi in generale H. Allcott, “Paternalism and energy efficiency: An overview”, in *Annual Review of Economics*, 8, 2016, p. 145; H. Allcott, M. Greenstone, “Is there an energy effciency gap?”, in *Journal of Economic Perspectives*, 26, 1, 2012, pp. 3-28, a pp. 3 e 5; H. Allcott, C. Knittel, *op. cit.*; J.M. Sallee et al., *op. cit.*; M.R. Busse et al., *op. cit.*

29. Vedi J.M. Sallee et al., *op. cit.*; M.R. Busse et al., *op. cit.*, p. 220.

30. Vedi H. Allcott, C. Knittel, *op. cit.*, pp. 33-34.

31. Vedi D. Duncan et al., “Most consumers don’t buy hybrids: Is rational choice a sufficient explanation?”, in *Journal of Benefit-Cost Analysis*, 10, 1, 2019, pp. 1-38.

32. *Ibidem*, p. 30.

33. Vedi J. Graham, J.B. Wiener, “Co-benefits, countervailing risks, and cost-benefit analysis”, Duke Law School Public Law & Legal Theory Series 2024-43, 2024, p. 19.

34. Vedi H. Allcott et al., “Regressive sin taxes, with an application to the optimal soda tax”, in *Quarterly Journal of Economics*, 135, 2019, p. 1557.

35. Vedi Y. Chen et al., *op. cit.*

36. Vedi l’eccellente analisi in I. Ayres, Q. Curtis, *op. cit.*

37. Vedi S. Bhargava et al., “Choose to lose: Health plan choices from a menu with dominated option”, in *Quarterly Journal of Economics*, 132, 2017, p. 1319.

38. Vedi R.H. Thaler, C.R. Sunstein, *Nudge*, *cit.*

39. Ovviamente, dovremmo definire razionalità e irrazionalità, per poterlo dire con certezza.

EPILOGO

PREVENIRE IL FURTO

Coercizione, menzogne, inganno e manipolazione sono concetti apparentati – quattro nemici dell'autonomia. Indicano tutti forme di violenza alla capacità delle persone di compiere scelte. Dei quattro, la coercizione è la più brutale. Cosa importante, le persone di solito sanno quando sono state sottoposte a coercizione. La manipolazione è la più insidiosa: non è detto che le persone sappiano di essere state manipolate.

Non esiste una definizione unitaria della manipolazione, che si presenta in molte forme diverse. Molte affermazioni o azioni, però, risultano manipolative perché non rispettano la capacità delle persone di compiere scelte riflessive e deliberative. Alcune forme di manipolazione sono relativamente innocue: per esempio quando un politico, un datore di lavoro, un medico o un cameriere utilizzano il tono della voce e le espressioni facciali per incoraggiare certe decisioni. Alcune forme di manipolazione sono più dannose: per esempio, quando un venditore induce con l'inganno le persone a concludere accordi con clausole nascoste che portano a oneri esorbitanti.

Se non riuscite a sopportare la manipolazione, probabilmente non avete senso dello humour. La manipolazione è una caratteristica pervasiva della vita umana. Può essere ridicola e divertente, almeno all'interno di relazioni di fiducia. Perfino quando non è né ridicola né divertente, non sempre dovrebbe essere proibita. Questo è il motivo per cui il sistema legale spesso è in grado di trattare menzogne e inganno, ma è molto più in difficoltà nel prendere di mira la manipolazione. Abbiamo visto che, se la legge proibisse la manipolazione in quanto tale, non darebbe alle persone un preavviso adeguato in merito a ciò che è proibito; il termine è troppo vago. Abbiamo visto anche che un divieto totale della manipolazione farebbe un unico

fascio di molte forme di condotta (o di cattiva condotta) che la legge non dovrebbe condannare; il termine è troppo ampio.

Comunque, la manipolazione è spesso una cosa orribile. Può infliggere danni gravi. Nelle sue forme più problematiche, mina l'autonomia delle persone. Non rispetta la dignità delle persone. Tratta le persone come mezzi e non come fini – come strumenti e stupidi, come oggetti e non come soggetti.[1] Da qui viene l'obiezione più sentita alla manipolazione: perché non sono stato trattato con rispetto?

L'obiezione welfarista alla manipolazione parte dall'affermazione che chi deve effettuare una scelta in genere sa che cosa sia nel suo migliore interesse. Conosciamo i nostri valori. Conosciamo i nostri gusti. Sappiamo che cosa ci piace. Quando le nostre scelte sono il risultato di una manipolazione, possono non promuovere il nostro stesso benessere, proprio perché non siamo stati messi nella condizione di riflettere e deliberare, per noi stessi, le variabili e i valori pertinenti. Questo è particolarmente probabile che sia vero se la manipolazione persegue interessi egoistici o motivazioni negative.

Dal punto di vista welfarista, la manipolazione è solo presuntivamente sfavorita. Un manipolatore benigno e informato potrebbe rendere la vita migliore, magari anche molto migliore, alle persone. Sotto ipotesi realistiche, la presunzione contro la manipolazione è giustificatamente forte, perché è improbabile che chi ci manipola sia benigno o informato.

Alcuni dei casi peggiori di manipolazione si basano sull'inganno: influenze occulte o nascoste sulle scelte delle persone, che compromettono la loro capacità di compiere scelte riflessive e deliberative. I deepfake possono essere manipolativi; spesso lo sono. I manipolatori utilizzano il pantano a loro beneficio; ostacolano il nostro cammino con l'attrito e nascondono aspetti importanti della situazione. La manipolazione può verificarsi anche quando i venditori sfruttano le influenze sociali, come nel problema Barbie. Quando le persone acquistano beni che desidererebbero non fossero mai esistiti, perdono tempo e denaro. La loro vita può peggiorare, e di parecchio. Anche la loro libertà è a rischio.

La manipolazione può essere una forma di furto. Le norme sociali dovrebbero opporvisi. Nei casi più gravi, dovrebbe opporvisi anche la legge.

1. F. Jongepier, J.W. Wieland, “Microtargeting people as mere means”, in F. Jongepier, M. Klenk (a cura di), *The Philosophy of Online Manipulation*, Routledge, London 2022, pp. 156-179.

RINGRAZIAMENTI

Sono molte le persone che devo ringraziare. Robert Dreeson, il mio fantastico editor, è stato una guida eccezionale e mi ha offerto suggerimenti preziosi in molte fasi. Quattro revisori anonimi hanno inviato commenti estremamente utili, da cui ho tratto grande vantaggio. Sarah Chalfant, la mia agente, è stata sempre di sostegno e guida. Jared Gaffe, Tate Harms e Tascha Shariari-Parsa sono stati superbi assistenti di ricerca.

Ho lavorato su questi temi per molti anni. Per discussioni e commenti preziosi in vari stadi, ringrazio Oren Bar-Gill, Anne Barnhill, Jonathan Baron, Tyler Cowen, Elizabeth Emens, George Loewenstein, Martha Minow, Sendhil Mullainathan, Martha Nussbaum, Eric Posner, Lucia Reisch, Eldar Shafir, Richard Thaler, Mark Tushnet, la scomparsa Edna Ullmann-Margalit e Adrian Vermeule. Un grazie speciale a Bar-Gill, Reisch, Thaler e Ullmann-Margalit per il lavoro comune e le molte discussioni che hanno ampiamente contribuito a dar forma a questo libro.

Grazie alla Harvard Law School per il prezioso sostegno. Grazie ai partecipanti a seminari al Queens College, Cambridge University, per i commenti e i suggerimenti, sempre eccellenti. Grazie infine alla Oxford University, e in particolare al suo Accelerator Program, per l'aiuto prezioso nelle fasi finali, essenziali e accelerate. In queste fasi finali, alcune delle ricerche pertinenti sono state condotte nell'ambito dello Accelerator Fellowship Program dello Oxford Institute for Ethics in AI.

Questo libro è dedicato a Sendhil Mullainathan, che ho già citato, perché è un amico e collega incredibile, perché è divertente, e perché la sua mente

è così viva che rende viva anche la mente di tutti gli altri.

Ho utilizzato in queste pagine vari saggi precedentemente pubblicati, fra cui “Fifty shades of manipulation”, in *Journal of Marketing Behavior*, 2015, p. 1; “Manipulation as theft”, in *Journal of European Public Policy*, 29, 2022, p. 1959; “Sludge Audits”, in *Behavioral Public Policy*, 6, 2020, p. 654; “The ethics of nudging”, in *Yale Journal of Regulation*, 32, 2025, p. 413; “Brave new world? Human welfare and paternalistic AI”, in *Theoretical Inquiries in Law*, 26, 2025. In tutti i casi, il riorientamento e le revisioni sono stati sostanziali. Ho esaminato alcuni degli stessi temi in *The Ethics of Influence*, Cambridge University Press, Cambridge 2015; ci sono sovrapposizioni, in qualche punto significative, fra quell’analisi e questa. Il mondo però è cambiato, e non poco, e su alcune questioni anche le mie idee sono cambiate, e non poco.

SCIENZA E IDEE

Ultimi volumi pubblicati

356. D.J. Chalmers, *Più realtà*
357. D.C. Dennett, *Coscienza*
358. P. Odifreddi, *A piccole dosi*
359. M. Galli, *Una banale influenza?*
360. N. deGrasse Tyson, D. Goldsmith, *Origini*
361. S. Rossi, D. Prattichizzo, *Il corpo artificiale*
362. P. Burke, *Ignoranza*
363. E. Selya, *Salvador Luria*
364. L. Susskind, A. Cabannes, *Relatività generale*
365. B. Larsson, *Essere o non essere umani*
366. J. LeDoux, *I quattro mondi dell'uomo*
367. S. Talamo, *Misurare la storia*
368. T. Sharot, C.R. Sunstein, *Guardate meglio*
369. L. Floridi, *Filosofia dell'informazione*
370. G. Ravasi, *Ero un blasfemo, un persecutore e un violento*
371. S. Carroll, *Spazio, tempo, movimento*
372. D.C. Dennett, *Pensandoci bene*
373. T.S. Kuhn, *L'incommensurabilità nella scienza*
374. P. Ferri, *Volare oltre il cielo*
375. G. Benenti, G. Casati, S. Montangero, *Il computer impossibile*
376. D. Quammen, *L'evoluzionista riluttante*
377. P. Odifreddi, *Incontri ravvicinati tra le due culture*
378. N. Carr, *Superbloom*
379. C.R. Sunstein, *Come diventare famosi*
380. M. Taddeo, *Codice di guerra*
381. R. Battiston, *Energia*
382. S. Carroll, *La fisica dei quanti*
383. P. Vineis, L. Savarino, *Come sopravvivere alla crisi ambientale*
384. A. Aspect, *Se Einstein avesse saputo*
385. S. Sloman, *Il prezzo della certezza*

This file was downloaded from Z-Library project

Your gateway to knowledge and culture. Accessible for everyone.

z-library.sk z-lib.gs z-lib.fm go-to-library.sk

Official Telegram channel

Z-Access

https://wikipedia.org/wiki/Z-Library